

Time to halt the gravy train

As European integration gathers pace, the notion that scientists must be rewarded with tax-free salaries and other perks for working at a 'foreign' international laboratory has become an anachronism.

Generous tax-free salaries topped up with non-residence allowances; remuneration to cover school fees; regular, paid trips home: these are among the privileges enjoyed by scientists at Europe's leading multinational laboratories. The labs' member states would like to trim these perks, but they have so far found this almost impossible to do.

The custom of extending the privileges of international civil servants to scientists was established by CERN, the European laboratory for particle physics near Geneva. CERN was founded in 1954 as a beacon of international cooperation in a continent only recently ravaged by war. Making it an attractive place to work was a high priority.

Over the next two decades, other 'Eurolabs' were set up along similar lines. The European Southern Observatory (ESO) in Garching, near Munich, was founded in 1962; the European Molecular Biology Laboratory (EMBL) in Heidelberg in 1974. Working in a foreign European country was then still an unusual and often daunting experience, so these labs adopted similarly generous salaries and benefits packages.

But as Europe moves ever closer to political union, and as national laboratories become increasingly international in their recruitment strategies, the landscape has shifted. Is it still appropriate to compensate European scientists so generously for living in some of the continent's most beautiful cities? Researchers whose careers have similarly taken them to another European country — but who don't happen to work at one of the elite Eurolabs — would surely answer no.

Europ perks

At CERN, a young, unmarried graduate can expect to earn 47,500 euros (US\$43,000); a married physicist with two children, a PhD and six years' experience might earn 63,000 euros; senior scientists typically command more than 80,000 euros. These salaries are all tax-free. At DESY in Hamburg, Germany's main high-energy physics laboratory, the equivalent salaries are 18,500 euros, 52,000 euros and 62,500 euros, respectively — from which up to 40% is deducted in tax and social security payments. Yet DESY seems to have no problem attracting foreign nationals: only one-third of its scientists are German.

Foreigners who work for Germany's Max Planck Society — who make up to 10% of its scientific workforce — have similar salary and benefits packages to those at DESY. And remember, researchers in Germany are among Europe's better paid: senior scientists working for the CNRS, France's national research agency, earn around 51,000 euros per year, before deductions for tax and social security payments. Again, around 10% of CNRS scientists are foreigners.

The advantages of working in a Eurolab do not end with tax-free salaries. For example, CERN pays school fees of some 10,000 euros per year per child for foreign scientists; the chosen schools can be in any of the labs' member states. Every two years, scientists from abroad get a paid trip home. Bizarrely, more than 200 of CERN's 2,600 staff also have diplomatic status.

CERN isn't even the most generous of the Eurolabs. Indeed, when the lab reviewed its salary and benefits packages in 1995 and 1999, it ranked only fifth out of seven among comparable European-level

facilities. After its reviews, CERN's governing council decided not to tinker with the system, arguing that generous salaries are necessary to continue attracting the best physicists against strong competition from industry. Over the past decade, however, the member states that bankroll other Eurolabs have begun to question the favourable terms and conditions that their scientists enjoy — but they have struggled to shift the status quo.

Freezes

ESO's council, for instance, has tried hard to freeze salaries and squeeze privileges. From the early 1990s it departed from its policy of awarding annual pay increases based on recommendations of the Coordinated Organizations, a committee of the Organisation for Economic Co-operation and Development that helps set pay for international bodies including the North Atlantic Treaty Organization.

Instead, salaries were kept down to, or below, inflation. But ESO's staff association complained to the International Labour Organization (ILO), and won its case. ESO's council agreed to compensate, but then changed its rules to give the council more leeway in deviating from the recommendations of the Coordinated Organizations. The ILO now argues that ESO has not clearly justified its salary decisions, and more cases are pending. ESO has, however, managed to check some privileges. For example, school-fee allowance has been frozen at some 7,200 euros per child per year for the past decade.

The EMBL has similarly fallen foul of the ILO in its attempts to control salary costs — and earlier this month was forced to agree to pay in full a claim for backdated implementation of higher salary scales that will cost the lab some 1.25 million euros.

Neither officials nor scientists at the Eurolabs like to talk about the issue of salaries and benefits. But they should, because an important principle needs to be addressed. Although it would be ludicrous to suggest that employment packages should be reduced to the lowest common denominator, or that they can be standardized across different types of laboratories, it must be right to question the silent continuation of perks that no longer have a clear practical or political justification.

This is particularly relevant in the current harsh economic climate, in which the Eurolabs may be forced to consider eating into their research budgets. The EMBL's governing council, for instance, will decide next March whether member states will stump up extra funds to pay for the backdated salary payments. If not, the money will have to be found from within the lab's existing budget.

CERN, meanwhile, is currently in considerable financial difficulties, facing a major overrun in the cost of building the Large Hadron Collider (LHC), its next big accelerator (see page 841). The lab's council is now reviewing the costs of all experiments that are peripheral to the LHC's main aim, the search for the Higgs boson, with a view to making savings.

Among the experiments under review are those that do not use the main accelerator's beamline, in which particle beams are fired at fixed targets. Their annual running cost amounts to some 10 million euros — about the same sum that CERN spends each year on non-salary benefits. ■





Moore's law
Intel founder is part
of Bush's science
advisory panel
p832



Fusion fears
Japan's plan to merge
nuclear institutes
fuels concern
p833



Final scores?
University research
audit may have given
its last grades
p834



Twin trouble
Two NASA
spacecraft start to
malfunction
p835

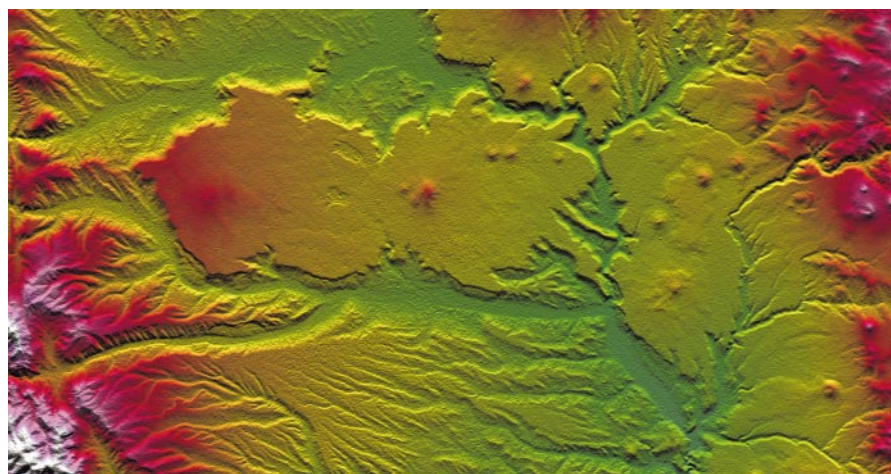
Map data kept under wraps as Pentagon focuses on security

Jonathan Knight, San Francisco

Geophysicists keen to access the most comprehensive topographical maps of the Earth ever produced are facing a longer wait than expected. The US Department of Defense has delayed the release of the maps, citing unspecified security concerns.

The maps are currently being assembled from data collected by the Shuttle Radar Topography Mission (SRTM). This joint project between the defence department and NASA — with additional contributions from the German Aerospace Center and the Italian Space Agency — aims to provide high-resolution maps of 80% of the Earth's land surface by the end of 2002.

One part of the project is mapping changes in elevation of as little as 8 metres for each continent. The first of these maps, covering North America, is scheduled to be finished in spring 2002. Another part of the effort is focusing on sites of scientific interest and is generating maps of 'cells' covering one degree longitude by one degree latitude. So far, 138 such cells



Clampdown: low-resolution images, such as this of Argentina, are available from the SRTM project.

have been created, and these were expected to be released as soon as they were finished.

But after the 11 September terrorist attacks, the Pentagon's National Imagery and Mapping Agency (NIMA), which co-

sponsors the project, placed a moratorium on any new releases of data pending a review of their implications for national security.

Officials at the agency will not say what the security implications of releasing the data might be. But accurate topographical data have various obvious military uses. They are indispensable, for example, for the navigation of low-flying, high-speed aircraft and of cruise missiles.

On 5 December, NIMA cleared the release of SRTM cell maps of the United States and its territories on the grounds that similar data are already in the public domain. These new maps are now available and join the handful of cell images of an area in Colorado that had been released before 11 September.

Data for the project were collected during a space-shuttle flight in February 2000. The Earth's surface between the latitudes 60 degrees north and 56 degrees south were scanned and, by comparing radar images taken from each end of a 60-metre boom, the SRTM's interferometer measured points of elevation on the Earth's surface. The measurements were originally expected to be within 16 metres, but researchers estimate that the instrument achieved twice that precision.

The bulk of the data are being processed by NASA's Jet Propulsion Laboratory in Pasadena, California, where a team is assembling a global map from the 12 terabytes

Spain sets sights on fusion facility

Sally Goodman

Spain is considering a late bid to host the proposed ITER international fusion research facility.

The Madrid-based Research Centre for Energy, Environment and Technology (CIEMAT) is carrying out a feasibility study into the prospect of a Spanish site for ITER. If a suitable site is identified and backed by the government, Spain would compete with France, Canada and Japan to host the facility.

"We see this as an important opportunity to develop high technology in Spain," says Felix Yndurain, CIEMAT's director general, "but we haven't yet discussed where the money will come from." He adds that Spain will decide in about three months whether to pursue the idea further.

At the beginning of 2002, Spain will assume the presidency of the European Union

— a platform that is often used by member states to launch grandiose initiatives. Spain may decide that an offer to host ITER would appropriately signal its growing scientific and technical aspirations. But critics point out that the country has no special expertise in the technology of magnetic fusion that will be used by ITER.

The United States' withdrawal from ITER in 1998, as well as the doubts of the other partners — Japan, Russia and the European Union — about the likely cost of the original ITER design proposal, have caused the project to be slowed and scaled back. The partners say that they are negotiating a legal framework for it, but will not choose a site until the end of next year at the earliest. Their choice will rest largely on which — if any — bid can muster the political support that will be needed to complete the project. ■

of information gathered.

At the annual meeting of the American Geophysical Union in San Francisco earlier this month, scientists who had applied more than a year ago for access to the emerging maps were forced to cancel a presentation that would have been based on information from the images.

"We had hoped that the restriction would be lifted in time for the meeting," says Peter Mougini-Mark, a volcanologist at the University of Hawaii in Honolulu who wants to use the cell maps to study the geological effects of the 1991 eruption of Mount Pinatubo.

Although most of the Earth has been mapped in detail from space, the elevation information for many areas is incomplete. Mougini-Mark and several other scientists who presented their SRTM plans at the meeting are still waiting for the non-US maps. "The whole community is very anxious to get them," says Mougini-Mark.

John LaBrecque, the NASA liaison with the Pentagon on the SRTM, says he has no idea when NIMA will release the non-US maps. "It is still under review," he says.

Even when the maps do enter the public domain, NIMA's original agreement with NASA allows the military to keep the most accurate versions for itself. For territories outside the United States, the public maps will have a resolution only one-third as precise as those held by the US military. ■

♦ www.jpl.nasa.gov/srtm

Bush turns to Silicon Valley moguls for scientific advice

Tony Reichhardt, Washington

The new President's Council of Advisors on Science and Technology (PCAST) — named by the White House last week — is dominated by computer-industry executives.

Bench scientists are thinly represented on the 24-strong panel, whose job it is to guide the White House on science and technology matters, but whose influence has varied between administrations.

Among the new PCAST members are two legendary figures from the information-technology industry: Michael Dell, chairman of Dell, and Gordon Moore, chairman emeritus of Intel. Robert Herbold, a vice-president at Microsoft, Carol Bartz, chief executive of software supplier Autodesk, and George Scalise, president of the Semiconductor Industry Association, are also on the panel.

In addition, President Bush has tapped some old hands from previous Republican administrations including Walter Massey and Bernadine Healy, who headed the National Science Foundation (NSF) and the National Institutes of Health, respectively, under his father. Another former NSF director, Erich Bloch, will be on the council, as will two holdovers from President Clinton's PCAST — Charles Vest, president of the Massachusetts

Institute of Technology, and retired aerospace executive Norman Augustine.

Newcomers on the panel include Charles Arntzen, a biotechnologist at Arizona State University, and Kathleen Behrens, a biotechnology analyst with investment-management firm Robertson Stephens. University administrators asked to join include Wayne Clough, president of the Georgia Institute of Technology, and Marye Anne Fox, chancellor of North Carolina State University.

PCAST will be co-chaired by White House science adviser John Marburger and Floyd Kvamme, a Silicon Valley venture capitalist.

Bush met with the group immediately after its first meeting on 12 December, which optimists take as an encouraging sign of his interest in the panel. White House officials say that PCAST will concentrate its early efforts in four areas: energy efficiency, information technology, counter-terrorism, and an overall assessment of federal research funding. They add that the heavy representation of information technology on the panel reflects its importance to the economy and to science.

According to a source familiar with Bush's plans for PCAST, the panel will concentrate on preparing memos to the president, rather than public reports. ■

Canadian budget cranks up investment in research

David Spurgeon, Montreal

The Canadian government has unveiled plans to spend Can\$7.4 billion (US\$4.7 billion) next year on science and technology, an increase of 8% over the previous year.

The announcement, which continues a major revival of Canadian research spending that has been under way since 1998, was made on 10 December as part of an overall budget that emphasizes large increases in military spending and counter-intelligence.

The budget includes an extra Can\$36.5 million for the Natural Sciences and Engineering Research Council (NSERC), which funds university research in engineering and the physical sciences, taking its budget to Can\$560 million. The National Research Council (NRC), which runs most of the federal government's research laboratories, gets a Can\$40-million increase, giving it a budget of Can\$536 million. The Canadian Institutes of Health Research will see an increase of Can\$75 million, lifting its budget to about Can\$560 million.

The NRC has been allocated Can\$110 million over three years to support recent

high-technology ventures such as a planned national institute for nanotechnology. The government will also provide a one-off payment of Can\$200 million to help universities to meet research overhead costs, and a further Can\$110 million to build a new, high-capacity broadband network to link research institutions and universities.

"Given the circumstances, I think it's quite a balanced approach," says John de la Mothe, a science-policy expert at the University of Ottawa. "Overall, it's good for the universities and very positive for the research community."

Bob Wolkow, a physicist at the NRC's Steacie Institute of Molecular Science in Ottawa, says he is "pretty pleased" with the budget. He adds that, given the government's fiscal difficulties, it is "all the more remarkable that we have got pretty substantial new funding".

But David Schindler, an ecologist at the University of Alberta who recently won the Herzberg medal for academic excellence from the NSERC, argues that the government should worry less about



Canadian finance minister Paul Martin (right) soaks up the applause for his budget speech.

international terrorism and more about pollution at home. "It just looks to me like they've got their priorities backwards," he says. Schindler adds that Canada's politicians do not seem to care about environmental damage, despite the fact that seven deaths have occurred from *Escherichia coli* contamination in Ontario in the past year. ■

EU ministers temper Framework reforms

Quirin Schiermeier, Munich

Radical plans for the European Commission's next Framework programme for scientific research have been watered down by European Union research ministers. The plans would have shifted money into continent-wide research networks and ended the funding of many small projects.

The commission's sixth Framework programme will distribute 17.5 billion euros (US\$15.8 billion) starting in 2003. Officials said in February that they hoped to use the funds to integrate European research activities by establishing networks in key scientific areas, such as functional genomics and nanotechnology. Average grant sizes were to increase by an order of magnitude, and support for smaller, unrelated projects would have ceased (see *Nature* **410**, 4; 2001).

The plans were criticized in November by the European Parliament, which proposed over 300 changes, including a shift in emphasis away from new fields such as genomics and towards diseases such as malaria and AIDS.

Meeting in Brussels on 10 December, research ministers from member states agreed to a compromise version of Framework 6 in which the large networks will receive slightly less funding than originally intended so that some smaller projects can



still be funded. Ministers also agreed that the new networks will be evaluated independently after two years.

The parliament had also wanted to curtail plans to support big computing projects such as the Grid — specialized software and high-speed network hardware that will link individual computers to create huge amounts of computing power (see *Nature* **414**, 475; 2001). CERN, the European laboratory for particle physics near Geneva, is one of several European organizations relying on the Grid to help them

analyse data from future projects.

Now, 300 million euros have been allocated to the Grid and related projects, close to the 350 million euros earmarked in the original budget. David Williams, coordinator of EU relations at CERN, says he is not complaining. "This is a perfectly sensible number which will help us get Grid networks and other e-science projects started."

The plans were also welcomed by representatives of central and south-eastern European countries, who hope to join the EU in the next few years, and fear that a big shift towards large projects would favour countries that already have a strong science base. "This is good news," says Andrzej Siemaszko, who runs a government office based in Warsaw that helps scientists write EU project proposals.

The new budget also features two new funding streams. In the first, 330 million euros will help to coordinate national research activities and policies — a cornerstone of the proposed common European research area. And 570 million euros are set aside for unforeseen developments that emerge during the Framework's lifetime, countering complaints about the inflexibility of the funding. ■

Planned merger worries Japan's nuclear researchers

David Cyranoski, Tokyo

Several hundred Japanese physicists and engineers face an uncertain future after the government announced plans to merge two nuclear research institutes by 2006. The move follows a decision earlier this year to cut the budgets of both institutes.

Details of the plan to unite the Japan Atomic Energy Research Institute (JAERI), based in Tokyo, and the Japan Nuclear Cycle Development Institute (JNC), located in Tokai northeast of the capital, were expected to be revealed on 18 December. Research at the JNC focuses on fast-breeder reactors and recycling of spent nuclear fuel. JAERI carries out a wide range of research, some of it related to atomic energy.

The proposal comes amid attempts by the government to streamline Japan's research institutes and to make research more economically focused. Plans to merge three major space laboratories were announced in August (see *Nature* **412**, 843; 2001).

The latest plan was given added impetus by the fact that both institutes are state-funded businesses known as 'special public corporations'. Many of these corporations have been criticized for their inefficiency, and

Prime Minister Junichiro Koizumi has vowed to change the status of them all by 2006. Government officials said earlier this year that budgets for JAERI and the JNC, like those of many other special public corporations, will be cut by 10% next year.

The Japanese science and education ministry, which helps to run both organizations, says that JAERI and the JNC have much in common, and that both institutes will benefit from a merger. Staff at the institutes disagree. "JAERI and JNC do completely different research and operate on a different scale," says Kazuo Mukai, director of planning at a JNC centre in Tsuruga, on the country's west coast. Shinzo Saito, vice-president of JAERI, adds that research at his institute could suffer from competition with JNC's costlier projects.

Both organizations are likely to be hit before the merger by the reduction in their budgets. Saito says that the cuts could affect JAERI's plans for a five-year project, worth ¥15 billion (US\$120 million), to build a prototype high-temperature, gas-cooled reactor. The reactor is designed to be more efficient than conventional reactors, and to produce hydrogen as well as electricity.

JAERI's construction of a linear accelerator and synchrotron at its Tokai campus could also be delayed. The facilities are due to be used in studies ranging from protein structure to superconductivity, and will form a booster for a larger synchrotron, which is scheduled to open in 2006.

Mukai says that it is too early to say how the cuts will affect the JNC, but predicts that many projects will be delayed by a year or more. ■



Man with a plan: Junichiro Koizumi wants to make state-backed enterprises more efficient.

TSUGUFUMI MATSUMOTO/AP

British research audit may be last of its kind

David Adam, London

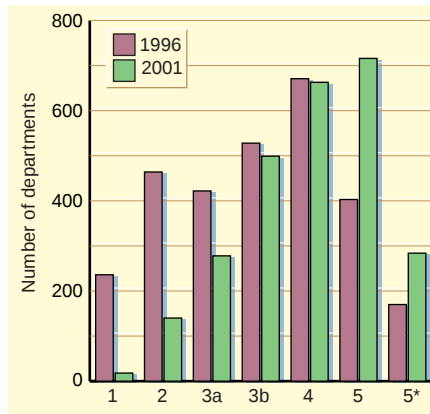
Researchers in the United Kingdom have an added reason for Christmas cheer — more than half of them now work in university departments that are rated as “world class”. These are the findings of the British government’s latest Research Assessment Exercise (RAE), a major audit of university research.

The audit’s results will be used by higher-education funding councils to distribute £1 billion (US\$1.5 billion) of government money each year for institutions to spend on salaries, buildings and equipment. Departments with the best grades get the lion’s share of the cash, whereas poor scorers get nothing.

However, while the many departments with improved grades celebrate the outcome, some are suggesting that the exercise should be the last of its kind. Critics argue that, with around 40% of departments scoring 5 or 5* — the best possible grades — and with others desperately skewing their priorities to join this enlarged elite, it is time to replace the exercise with something more subtle.

“The introduction of the RAE was a good idea,” says Richard Sykes, head of Imperial College in London. “It made people sit up, and it has certainly improved standards. But I think it has lived its life now. I would like to see this being the last RAE of the old style.”

Since its introduction in 1986, the RAE has attracted widespread interest from other governments around the world, most of which would like to make their research universities more efficient. This year, the exercise itself cost some £40 million, and used 60 separate sub-



Grade inflation: department scores have surged strongly upwards since the last audit, in 1996.

ject panels to assess around 200,000 published articles from almost 50,000 researchers. The panels were asked to judge each article’s level of excellence — whether national or international — using a standard scale.

This year’s results, released on 14 December, showed a strong increase in the number of university departments carrying out research rated as internationally excellent: some 55% of researchers now work in such departments, compared with 31% in 1996.

University and government officials insist that the increase in the quality of the work assessed is genuine. “The improvements in performance since the last RAE are a direct result of institutions managing their research strategically,” says Howard Newby, chief executive of the Higher Education Funding Coun-

cil for England (HEFCE). “Universities and colleges are to be congratulated on this outstanding performance.” According to Newby, the world-class reputation of departments given top grades has been confirmed by some 300 international experts, and publication and citation counts also suggest that the quality of UK university research is improving.

In each of the five RAEs conducted since 1986, the grades of the 1,000 or so departments surveyed have steadily improved. But there is no doubt that department heads have learned to play the system. Some, for example, have dissuaded researchers from spending time writing books, as these do not count in the RAE’s reckoning. “We know the RAE has some distorting effects,” says Peter Cotgreave, who heads the pressure group Save British Science.

Others point out that, by focusing on the strengths of single university departments, the exercise fails to encourage interdisciplinary research — and may even cause researchers to concentrate on single-discipline projects that fall neatly into an established RAE subject category.

The number of departments now scoring high grades is also causing a headache for the HEFCE and the other funding councils. As the Dodo put it in *Alice in Wonderland*: “Everybody has won, and all must have prizes.” The HEFCE has said that if the money was awarded using the same funding formula as in 1996, there would be a £200-million shortfall. And it has said that it will have to cap large budget increases owed to universities that have improved their standing.

Cotgreave says that the rise in the number of departments winning the best grades will make similar future tests less helpful. “The RAE has now eliminated much of its ability to discriminate,” he says. “If all that will happen is that you discover that everybody is equally good, then it is pointless doing it again.”

But if this RAE is the last of its type, what will replace it? “As long as we have selective funding of research we must have some kind of research assessment exercise,” says John Rogers, RAE manager for the HEFCE. “But we should question whether we need to continue with this big-bang exercise every five years. Perhaps we could have a rolling programme or something with a lighter touch.”

Alan Wilson, vice-chancellor of the University of Leeds and chairman of the research group of the umbrella organization Universities UK, suggests two possibilities. One would be a smaller sampling audit of departments. The second would be for the higher-education councils simply to award money in direct proportion to departments’ funding from the research councils. “My personal view is that there has to be some kind of evaluation, but that it could be much simpler,” Wilson says. ■



Time for a change: Imperial College’s Richard Sykes says that the exercise has run its useful course.



Indian government orders review of human drug trials

New Delhi The Indian government has ordered an immediate review of the ethics and safety of all human drug trials being conducted in the country.

The move was prompted by controversy surrounding the trial of a cancer drug at the Regional Cancer Centre in the southern Indian state of Kerala. Researchers from Johns Hopkins University in Baltimore, Maryland, conducted human trials of a drug that had not been properly approved by the appropriate bodies in either India or the United States (see *Nature* **412**, 466; 2001).

The health ministry says that India will not be used as a low-cost testing ground for drugs developed abroad, and that action will be taken against any researchers and institutions that contravene the ministry's guidelines on biomedical research. All human trials at the Regional Cancer Centre have been suspended for six months, and staff involved in the trial are likely to face disciplinary action.

Plasma physicist welcomed as US energy science chief

Washington President George W. Bush has announced his intention to nominate plasma physicist Raymond Orbach to be the next director of the Department of Energy's Office of Science.

Orbach, who is currently chancellor of the University of California, Riverside, will take charge of most of the department's research that is not related to nuclear weapons. This includes the main US physics research programmes, as well as the operation of its largest research facilities, such as the Spallation Neutron Source being built at the

Oak Ridge National Laboratory in Tennessee.

Orbach's nomination was greeted enthusiastically by leaders of the physics community, and is likely to be quickly confirmed by the US Senate.

Swedish advisers back therapeutic cloning

Stockholm Therapeutic cloning should be made legal in Sweden, say the authors of a government-commissioned study. The report, compiled by the Swedish Research Council, says therapeutic cloning is "ethically defensible" as long as it provides significant medical benefit.

Sweden does not currently have therapeutic cloning legislation in place, although it is a signatory to the bioethics convention of the Council of Europe, which bans all forms of human cloning for research purposes. The research council suggests that Sweden should now apply for a certificate of exemption from the convention. The authors also recommend permitting the extraction of stem cells from embryos left over from *in vitro* fertilization.

♦ www.vr.se

European strategists examine anti-terror options

Paris A group of scientific experts in charge of developing a strategy for countering the threat of biological and chemical weapons met for the first time on 13 December.

The group, which was established by the European Commission in response to the recent terrorist attacks in the United States, will assess the current state of research that could be used to tackle terrorism, such as studies of toxic agents and the development of surveillance systems. Research into vaccines and treatment strategies appropriate for biological weapons will also be examined.

Members aim to identify gaps where more research is needed, and will report back to European research ministers in March.

Equipment failure derails space projects

Washington Two NASA spacecraft have run into trouble. The system used to adjust the orientation of the Far Ultraviolet Spectroscopic Explorer (FUSE), which operates from an Earth orbit, shut down on 10 December. And engineers working with Cassini, which is on its way to Saturn, have reported problems with the craft's camera.

FUSE is two years into a three-year project to study hydrogen–deuterium ratios in interstellar space. Two of the four wheels that are used to adjust the orientation of the craft are currently malfunctioning. Staff from the Orbital Sciences Corporation, the Virginia-



Out of focus: Cassini's camera has developed a fault and is now sending back hazy images.

based company that built FUSE, are now working to devise an alternative means of adjusting the craft's position should they fail to fix the wheels.

Engineers working on Cassini, a joint project between NASA and European partners, have found haze on its recent images, which they suspect has been caused by a deposit on the camera's optics or light-detecting device. They hope to remove the contamination by heating the camera, and report that a first warming cycle seems to have been partially successful.

♦ www.jpl.nasa.gov/cassini

♦ fuse.pha.jhu.edu

Germany sees increase in laboratory animal deaths

Munich The rising use of genetically modified mice in biomedical studies contributed to a 15% increase in the number of animals killed in Germany last year for research. The figures, released last month, show that 1.83 million animals were killed in 2000, up from 1.59 million in the previous year.

Mice were the biggest sufferers — 976,000 were killed last year, an increase of 26% on 1999. Ivar Aune, a spokesman for the Society for Health and Research, Germany's largest lobby group for biomedical research, says that this is partly due to the increased use of transgenic mice. Changes in counting methods, which for the first time included animals killed for teaching purposes or for their organs, also contributed to the rise.

The trend is similar in Britain, where transgenic animals last year accounted for one-third of the 2.7 million animals killed for research. Research in both countries has been the subject of high-profile campaigns by animal-welfare activists.

The German government's figures also show that in the pharmaceutical and cosmetics industries fewer animals were killed as a result of a greater use of alternative methods.



Post pending: Raymond Orbach is poised to take charge of US physics research programmes.

2001 in context

It should have been a year of celebration, as science passed a major milestone with the publication of the human genome sequence. Instead, a global economic slowdown and the world-changing events of 11 September have cast a long shadow. *Nature's* reporters look back on a tumultuous year.

The big picture

Science in a changed world

On a calm and gloriously sunny September morning in Bethesda, Maryland, thousands of researchers were streaming into work as usual on the main campus of the National Institutes of Health (NIH). Among them were officials and reporters attending a meeting of the National Cancer Institute's advisory board, at which Richard Klausner, its director, was to explain why he was moving on to head a new health sciences institute.

As Klausner stood up to speak, an aide handed him a piece of paper. Klausner haltingly informed the audience that the World Trade Center was ablaze, and the meeting disintegrated as screens were switched to live television coverage of the attacks in New York and at the Pentagon, just a few miles away. For the people in that room, as for so many others in the United States and beyond, the world had suddenly become a frightening and much more uncertain place.

Three months later, researchers are struggling to comprehend how their lives will be affected by the ramifications of the events of 11 September. Scientists worldwide are also concerned about a global economic downturn that bodes ill for some of their sources of financial support. Twelve months ago, when *Nature* presented its overview of 2000, researchers were buoyant, celebrating a host of sequenced genomes and enthused by emerging new fields of study. But although scientific progress continues apace, the dominant mood is now one of anxiety.

Not surprisingly, the biggest changes have come in the United States. Take a visit to the NIH campus today, and the transformation is apparent. The relaxed, collegial atmosphere

has been fractured, and the site is ring-fenced with security. The NIH's mission has also shifted subtly, as its leaders ponder how they should respond to the threat of bioterrorism.

Much the same applies, to varying degrees, at every other US research agency. Between the end of the cold war and the attacks on 11 September, public health and economic competitiveness had established themselves as the predominant political justifications for the US government's vast investment in science. Now national security is back with a vengeance. The Department of Energy's weapons laboratories, for instance, have for years lurched from one crisis to another. But now they are finding that many of their security-related projects command intense public interest and support.

Even the National Science Foundation (NSF), which has the task of supporting 'pure' science, has been forced to consider the relevance of its programmes to the revised agenda. To researchers' relief, fears that the NSF's budget would be cut to release money for more immediate priorities are now receding. But although the NSF's centrepiece initiative — a push in mathematics — still seems to be on track, its rationale has been recast to include applications in intelligence.

The US scientific establishment is confronting these changes. On 26 September, a closed-door meeting at the National Academy of Sciences brought prominent scientists together with the likes of Sam Nunn, a former senator with expertise in national security, and James Woolsey, former director of the Central Intelligence Agency, to consider how science could assist the 'war on terrorism'.

In other countries, this conflict has had

less immediate implications for research. Scientists with appropriate expertise are being tapped for advice. And, as in the United States, investigations into visiting researchers from 'sensitive' countries are raising concerns about civil liberties and the freedom of scientific exchange. But across Europe and the Asia-Pacific region, there has been no fundamental shift in the research agenda. Indeed, for many of the world's scientists, worries about the current economic slump are more pressing than any need to justify their work in terms of national security.

The downturn hit first and hardest in the information and communications technology sector. Some companies remain optimistic about maintaining their research spending in the face of these difficulties, but others are in deep crisis — which is bad news for researchers working at the interface of physics and high-tech industry.

As the wider global economy slows, few scientists can afford to be complacent, whether they work in industry or academia. Although research budgets are, for the most part, not yet being cut, the largesse that science has enjoyed in many countries in recent years might now be curtailed. Future spending may also have to be justified more explicitly in terms of its potential to enhance economic competitiveness. Nowhere are these concerns more acute than in Japan. To the alarm of many scientists, its government has developed policy guidelines that place a clear priority on pushing scientists to underpin Japanese industry.

No one can tell for sure what the future holds, but at least some aspects of scientific life are returning to normal. In the immediate aftermath of 11 September, attendance at scientific meetings plummeted as researchers shunned air travel. But by the year's end, events such as the American Society for Cell Biology's meeting in Washington and that of



S. PLATT/GETTY



A. WONG/GETTY

The events of 11 September shook the world and put a question mark over scientific priorities — an issue further complicated by growing fears of bioterrorism, fuelled by anthrax attacks in the United States.

the American Geophysical Union in San Francisco were fully subscribed.

There will be no rapid return to the optimism that prevailed among scientists before that fateful September morning. But

as we move into 2002, the usual New Year's wishes for peace and prosperity will be made more fervently than for many decades past.

Colin Macilwain

even talk of barring the publication of results that could be applied to bioweapons development.

This debate is alarming many biologists. Rather than imposing restrictions on research, some experts argue that mounting an effective response to the bioweapons threat means harnessing the skills of the biomedical research community to develop countermeasures. "Our best defence is to have a lot of terrific scientists working on the problem," says Steven Block, a biophysicist at Stanford University in California.

Block, who is a member of the JASON group that provides advice to the US government on the technical aspects of defence and other matters, warns that vaccine technology lags behind other areas of molecular medicine. A major investment in vaccine research would be an important first step, he says.

There is also a need for a better understanding of the epidemiology of bioweapons agents. US public-health officials first assumed that the anthrax letters could not have infected anyone until they were opened. But this proved fatally wrong when two postal workers died of inhalation anthrax. Later, the deaths of Kathy Nguyen, a New York City hospital worker, and Otilie Lundgren, a 94-year-old Connecticut woman, caused confusion, because neither had been near the tainted letters.

With a better appreciation of how weaponized anthrax spreads, both events might have been anticipated. Matthew Meselson, a molecular biologist at Harvard University who in 1979 led an investigation of the accidental release of anthrax from a Soviet bioweapons facility in Sverdlovsk, believes that the conventional wisdom that a threshold number of spores must be inhaled before disease can occur is wrong. Even a single spore might cause disease, albeit

Bioweapons

Delivering death in the mail

We can't claim that we weren't warned. For years, experts have argued that most nations are ill-prepared to deal with a bioterrorist attack. Events of the past few months have confirmed their assessment in chilling fashion.

Thankfully, whoever was responsible for sending anthrax in the mail to US politicians and media organizations either lacked the means, or the will, to release an aerosol of spores over a major city or in a subway system. But despite the low-tech means of delivery, these were the first bioterrorist attacks to claim human lives.

If readiness is measured in terms of vaccine supplies, the situation is dire. The current vaccine for anthrax requires six injections over two years to induce immunity. Present US stocks are all spoken for by the military, and the vaccine's sole producer, BioPort of Lansing, Michigan, has not shipped a single dose since 1998 — when it failed to meet government quality-control standards.

For smallpox, the US government has maintained a stock of 15.4 million doses of dried vaccine that it hopes can be diluted to give at least 77 million doses. But many experts believe that this is not enough.

Since October, the United States has taken steps to bolster vaccine reserves. Researchers

at the Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland, accelerated their efforts to get a newly developed anthrax vaccine into clinical trials. In November, the Department of Health and Human Services announced a \$428-million order for 155 million doses of smallpox vaccine from Acambis, a company in Cambridge, UK — in addition to 54 million doses that were already on order.

Suddenly, biodefence has become a priority. Less than two weeks after the death on 5 October of Robert Stevens — who inhaled spores from a letter posted to a tabloid publisher in Boca Raton, Florida — President George W. Bush requested \$1.5 billion to purchase vaccines and antibiotics, to bolster the public-health system and to develop surveillance and detection technologies. Next year's budget request may see further calls on the US public purse.

What this means for scientists at the bench remains unclear. It is possible that some of the additional funds will be devoted to research. But some politicians and their advisers have suggested reining in the open culture of bioscience because it offers easy access to data and materials that can be used for malevolent purposes. New restrictions on handling dangerous pathogens have already been passed into law in the United States, and tighter controls may follow soon. There is



Raw potential: embryonic stem cells, shown preparing to differentiate (left), came under intense scrutiny this year as researchers struggled to gain permission to investigate their therapeutic power.



D. SILVERMAN/GETTY

infrequently, he argues. So Nguyen and Lundgren may have been the unlucky victims of low-level anthrax contamination of the mail by the tainted letters.

Although anthrax and smallpox are worrying, a well-funded bioweapons programme could engineer pathogens to be even more deadly. In the long term, Block forecasts a “molecular arms race” between bioweapons developers and biodefence specialists.

To many experts, this prospect underlines

the need for an effective mechanism to enforce the 1972 Biological Weapons Convention, the international agreement that is supposed to outlaw bioweapons production. But attempts to give the treaty teeth founded this year, after the United States rejected the proposed monitoring scheme as ineffective and likely to compromise the commercial secrets of its biotechnology industry. Subsequent events have not changed the Bush administration's mind.

Jonathan Knight

there was an important caveat: the cell lines involved must have been derived before the date of his announcement.

Biologists don't know whether access to these lines will be sufficient to reap the potential of ES cells. Many of the 60 or so approved cell lines have not been fully characterized, and it could be that lines that have yet to be isolated will emerge as the preferred materials for researchers in the field.

More generally, proponents of regenerative medicine are unsure whether ES cells will remain the field's brightest stars. Alternatively, it might eventually become possible to coax the ‘adult’ stem cells that reside in many of our tissues to achieve similar results.

But the idea of ‘therapeutic cloning’ seems to be on the wane. By creating cloned human blastocysts, some experts have argued that it should be possible to derive ES cells perfectly matched to individual patients. But most now believe this will be too expensive and cumbersome for regular clinical use.

Advanced Cell Technology (ACT), a biotech company in Worcester, Massachusetts, begs to differ, and gained publicity in November by announcing that it had created cloned human embryos. It claimed to have passed a milestone on the road to therapeutic cloning. But in fact, the best ACT had achieved was a six-cell embryo — far from a blastocyst. If anything, the results confirmed just how difficult therapeutic cloning will be.

In any case, research into therapeutic cloning may be legislated into abeyance in many countries. The US Senate is poised to pass a bill that would ban all forms of human cloning, both reproductive and therapeutic. And the practice is already effectively prohibited by the Council of Europe's bioethics convention — to which many European countries have signed up — which bans the creation of embryos for research.

Peter Aldhous

D. SILVERMAN/GETTY

Regenerative medicine

A world of difference

Location, location, location — for researchers exploring the therapeutic potential of human embryonic stem (ES) cells, the mantra of real-estate sales has acquired a fresh significance. ES cells hold the prospect of growing replacement tissues for those lost to disease, injury or ageing. But in this field, what you can do depends on where in the world your lab is based.

Among the major scientific nations, Britain emerged in 2001 as an enthusiastic supporter of ES-cell research, amending its law to allow ES cells to be isolated from human blastocysts — the hollow ball of cells that forms after some five days of embryological development — for research into regenerative medicine. Japan has also prepared guidelines that will give its researchers similar freedom. Sweden, Israel and Australia are also supportive of work on human ES cells.

But in several other countries, such work remains banned, or at least restricted. In France, for example, stem-cell researchers

have been campaigning for more than two years to amend the law to allow research on human ES cells, and their isolation from blastocysts. This is not now expected to happen until early next year, and the revised law may not come into effect until 2003.

In Germany, meanwhile, hostility to the field in some quarters is so great that scientists wanting to import human ES cells have received death threats. Deriving human ES-cell lines is prohibited by Germany's strict embryo-protection law, but a loophole means that importing the cells has not yet been banned. Although the DFG, Germany's main research granting body, has approved funding for human ES-cell projects in principle, it is currently sitting on its hands, waiting for the parliamentary and public debate to reach a conclusion.

In the powerhouse of biomedical research — the United States — confusion reigns. In August, President George W. Bush announced that federal funds would be released for research on human ES cells. But



Gaining ground: for many US scientists, George W. Bush's presidency got off to a poor start, but his handling of events since 11 September has brought him wider support and respect.



president's party control of the upper chamber of Congress. But it took the events of

11 September to forge better relations with the scientific community.

Bush's actions since then have increased his standing with the American public — scientists included. And researchers are pleased that physicist John Marburger has finally been confirmed as Bush's science adviser.

But the relationship between the research community and the Bush administration remains uneasy. Space scientists, for instance, are troubled by the appointment of a NASA administrator whose overriding priority is budgetary control (see page 841). For biomedical researchers, meanwhile, the continuing leadership void at the NIH remains a source of concern.

Colin Macilwain

US science policy

Under new management

Few relationships are more central to the progress of science than that between a US president and the nation's researchers. Rarely has this relationship got off to such a bad start as during George W. Bush's first few months in the White House.

What annoyed many researchers was Bush's failure to fill the top scientific positions in his administration, combined with his determination to push ahead with a series of policy moves that conflicted with the perceived scientific consensus.

The Bush administration's slow start at filling hundreds of key positions across the government came about because of the initial delay in determining the outcome of last year's presidential election. But as the months rolled by, the failure to nominate a presidential science adviser, on paper the most important scientific position in the US government, began to cause disquiet.

Of the other leading science posts, at the start of 2001 only one was occupied by someone who wanted to stay in the government — Rita Colwell, director of the National Science Foundation. Dan Goldin, the administrator of NASA, was prepared to remain for a short period only; the directorship of the National Institutes of Health (NIH) had been vacant since 1999.

In this administrative and advisory vacuum, Bush adopted controversial policies on a series of issues in which science has an important stake, including climate change and ballistic-missile defence.

The decision in March not to ratify the Kyoto Protocol, the international treaty to combat global warming, was not a complete

surprise. But the absence of consultation, and the failure to develop alternative approaches to the problem angered many researchers — particularly coming in the wake of a budget request that proposed deep cuts in environmental science.

Ballistic-missile defence, meanwhile, quickly moved to the top of the Bush administration's foreign-policy agenda — despite concerns raised by physicists about the plan's feasibility, and the wisdom of abandoning the Anti-Ballistic Missile Treaty to prepare for its deployment.

Bush's bullish approach was tempered somewhat by the defection in May of a Republican senator — a move that lost the

Animal health

The killing fields

British farmers have grown used to tough times. But nothing could have prepared them for an epidemic of foot-and-mouth disease (FMD) that took a cull of almost 4 million animals to eradicate.

The scientists who dealt with the outbreak are unlikely to forget this year either. Epidemiologists and veterinary scientists were thrust onto the public stage — and sometimes into conflict with one another — as supporters of a tough culling policy clashed with advocates of vaccination.

The strain of FMD virus involved, known as pan-Asian type O, was first spotted in India in 1990. From there it spread through Asia and the Middle East, displacing existing strains in

many countries. Lacking the resources to vaccinate or cull animals to stamp out the disease, farmers in the developing world have been forced to live with the virus, which reduces meat and milk production.

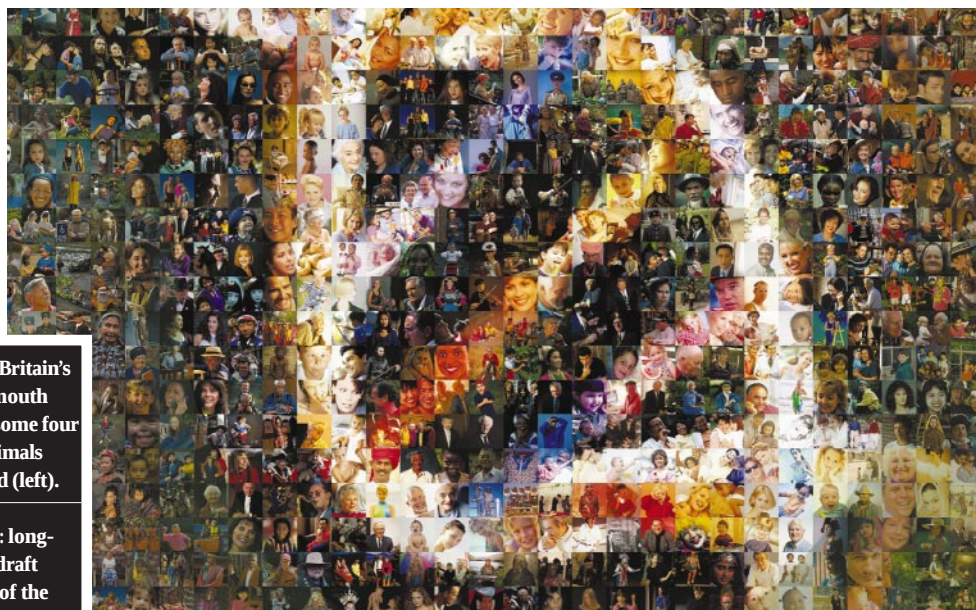
But Britain had been almost entirely free of FMD for some 30 years. Many of the country's trading partners will not accept meat or animal exports from countries with FMD. So the aim was to eradicate the virus as quickly as possible.

The first British cases were confirmed on 20 February, and country-wide restrictions on animal movements were imposed the following day. Unfortunately, infected sheep had already been moved through a market in



Dead loss: Britain's foot-and-mouth crisis saw some four million animals slaughtered (left).

Base camp: long-promised draft sequences of the human genome finally appeared in print (right).



S. TOUCHIG/NEWSMAKERS

northern England, and these animals spread the disease across the country. Just over a month later, the number of infected farms had topped 700.

As the outbreak spread, it became clear that the Ministry of Agriculture, Fisheries and Food (MAFF) was unprepared for such a calamity. Understaffing left local vets struggling to keep up with new cases, and MAFF's plans for responding to emerging data were woefully inadequate — many local veterinary offices were not even on e-mail. Indeed, MAFF never recovered from criticism of its handling of the crisis, and has since been reconstituted as the Department for Environment, Food and Rural Affairs.

In early March, three teams of independent epidemiologists started crunching the data. They forecast a 'meltdown' unless a much more effective culling policy was introduced. On hearing this message, the government's chief scientific adviser, David King, wrested control of the crisis from MAFF and formulated a policy that leant heavily on the epidemiologists' mathematical models.

The plan was to kill infected animals within 24 hours and those on surrounding farms within 48. It worked: the last case — on farm number 2,030 — was recorded on 30 September. But was the cost too high? Some veterinary scientists feel that vaccination would have limited the epidemic's size, and complain that King paid little attention to their views. They cite the use of vaccination elsewhere, including a programme used this summer in the Netherlands to quell a nascent foot-and-mouth outbreak sparked by imports of infected animals from Britain.

The epidemiologists who advised King argue that their models show that the large size of the epidemic — and hence the huge cull needed to eradicate it — can be blamed on the initial delay in reducing the virus's spread.

Epidemiologists argue that retrospective

analyses of the British outbreak confirm that vaccination would have had little effect. They claim

that any benefits from vaccination would have been outweighed by the adverse effects of moving resources away from the cull programme, and note that vaccination does not save animals from slaughter. Indeed, vaccinated livestock must be killed if exports are swiftly to be resumed, as simple tests cannot distinguish immunized from infected animals. But critics reply that the full range of

vaccination strategies has yet to be modelled.

Clearly, more work is needed before there is a consensus on the strategy to be used against future outbreaks. But the wounds opened up by this year's British epidemic may mean that considerations other than the output of mathematical models will come into play. "Culling on such a scale will not be politically acceptable," says Mark Woolhouse, a veterinary epidemiologist at the University of Edinburgh, and one of King's advisers. ■

Jim Giles

E. LANDER/WHITEHEAD INSTITUTE

Postgenomics

Data, data, everywhere...

The full ramifications may not yet have sunk in for many biologists, but a revolution is under way. For those comfortable with postgenomic techniques, a galaxy of scientific opportunities beckons.

To take advantage of this, many branches of biology may need to be transformed — moving from bench sciences dominated by competing research groups into collaborative enterprises. This would see entire research communities organized around databases and associated computational tools. In such fields, these online resources could eventually replace the conventional scientific paper as the predominant form of communication.

Although this year saw some early stages in this transition, there was first some unfinished business from 2000 — the year of genome sequences — to attend to. After months of anticipation, rival versions of the draft human genome, produced by the international Human Genome Project and Celera Genomics of Rockville, Maryland, were finally published in February.

As researchers attempt to link the sequence data with information on gene function and expression, protein production and even clinical records, a plethora of techniques are being used. From DNA microarrays for studying gene expression to advanced techniques for protein analysis, the common thread is the generation of unprecedented quantities of data: we are entering the era of high-throughput biology.

Increasingly, researchers will be able to apply sophisticated computational tools to huge data sets to model the behaviour and function of whole pathways of genes and proteins. And as more genomes become available, computational approaches will also be key to gaining insights into gene function through comparative studies.

Most biologists will need assistance if they are to get a piece of this action. But thankfully, experts are trying to help. The international Microarray Gene Expression Database Group, for instance, is developing common protocols for high-throughput



In the red: costs forced NASA to revise plans for the International Space Station (left), while problems with superconducting magnets (right) helped to push CERN's Large Hadron Collider over budget.



gene-expression studies, from the preparation of samples to the visualization of microarray data. Biologists new to the technology will be able to plan their research using these standards, enter their results into a common global database, and use its software tools to query the pooled data sets.

Meanwhile, human-genome browsers such as Ensembl, run from the European Bioinformatics Institute in Hinxton, near Cambridge, or that hosted by the University of California, Santa Cruz, already give access to layers of comparative genomic, functional genomic, genetic, cytogenetic and clinical data all anchored to the underlying sequence data.

This year saw a lively debate over online publishing in biology, as a grassroots initiative called the Public Library of Science tried to convince publishers to deposit papers in a free-access online database six months after their initial publication. But for a vision of the future of biological communication, speak to those researchers who are most enthusiastically embracing the high-throughput approach. The Alliance for Cellular Signaling, for instance, which involves researchers from 20 institutions across the United States, sees its planned database and associated web pages as the primary means of disseminating its results. Conventional scientific papers simply cannot do the job, argue its leaders.

Declan Butler

problems with financial management are resolved. But ultimately the Young committee called for a fully capable station with a crew of six, clearly dedicated to science.

CERN's problems are not in the same league as NASA's, although its management has faced harsh criticism. Rumours that the LHC project was struggling to keep to its austere budget began circulating in the spring. After months of denials, CERN finally admitted in September that the project was likely to overspend. Its managers pointed to a lack of contingency funds, problems with the development of superconducting magnets and escalating civil-engineering costs; critics blamed financial mismanagement. Either way, CERN's member states are in no mood to bail out the project.

CERN director-general Luciano Maiani remains optimistic that the LHC will still open on time. But how this will be achieved has yet to be determined, and strategic control of the project has effectively been handed to an external review board.

As NASA and CERN attempt to balance their books, researchers working on similarly ambitious projects are watching anxiously. In the current economic climate, they realize, talk of spiralling costs may make politicians nervous about committing money.

High-energy physicists trying to sell the next big machine after the LHC — an enormous electron-positron collider — are likely to face some awkward questions. The same goes for astronomers planning the Next Generation Space Telescope, scheduled for launch in December 2008. Its technical specification is ambitious, jumping from the 2.4-metre mirror of the Hubble Space Telescope to a proposed 8-metre reflector. Some of the project's proponents are already wondering if it might be wise to adopt a more modest plan.

David Adam and Tony Reichhardt

Big science

Down to Earth with a bump

For NASA, 2001 was the year in which the space odyssey that is the International Space Station ran into serious trouble. The script is hastily being rewritten, with the goal of controlling this blockbuster's budget.

The space station is the ultimate big-ticket project. But it was not the only such endeavour to run into serious budgetary difficulties this year. At CERN, the European high-energy physics laboratory near Geneva, a projected overspend of several hundred million dollars has left questions hanging over its next big particle accelerator, the Large Hadron Collider (LHC), which is supposed to start collecting data in 2006.

Few would dispute the judgement of incoming NASA administrator Sean O'Keefe that the space station is in "management and

financial crisis". Early in the year, NASA shocked the new US administration by saying that it would need \$5 billion more than the \$8 billion that had previously been budgeted to finish the project. In being handed over to O'Keefe, a budget and management expert, the space agency resembles a corporation that has gone into receivership. Financial responsibility, not technological and scientific vision, is the new priority.

Even before O'Keefe's nomination, the decision had been taken to freeze the station at an interim stage, with a maintenance crew of three astronauts on board, until costs could be contained. A committee appointed to look into the project, headed by veteran aerospace executive Thomas Young, agreed in November that NASA should stay with a stripped-down station for two years, until

Beware the baited hook of publicity

Headlines are tempting but lead to disillusionment — and a sour taste with your spaghetti.

Sir — In his Words essay “Good news is no news” (*Nature* **413**, 113; 2001), John Emsley suggests that scientists who want their work featured in the media should learn how to package their discoveries by finding the right “hooks” for the public.

Scientists have various reasons for wanting to make headlines. Informing the public about the state of a scientific field or of a discovery; attracting funds to hospitals and universities; advertising companies with which they are involved; or (all-too-common) self-aggrandizement — these are all valid reasons for talking about science. But they all should have the same common denominator: the truth.

Unfortunately, the media often embellish, sensationalize and exaggerate reality, which creates false illusions. I would advise caution in encouraging such behaviour. The danger is that a journalistic mentality could infiltrate even the scientific literature, as even prestigious journals can be guilty of publishing splashy and ‘sexy’ studies that are not scientifically sound.

My personal experience with the ‘tabloidization’ of scientific information began when I moved to Manhattan, and discovered some of the best Italian restaurants in the world. What does this have to do with science and media? A lot for me, as I have made friends with many Italian restaurateurs. Frequently, in the past few years, I have been approached by an Antonio, Giovanni or Marcello with “Have you seen it in the news?”. “No, what?” I answer. “An Italian doctor has discovered a cure for cancer (or this or that),” is the reply.

It was in 1998 that I first learned that Luigi Di Bella, an 85-year-old practitioner in a small Italian town, had discovered a miraculous cure for cancer (see *Nature* **391**, 217; 1998). The media had blown this up so much that the Italian parliament was forced to approve a clinical trial (see *Nature* **392**, 421, 1998) to test the ‘cure’: a cocktail of somatostatin and vitamins. After months of testing, there was no evidence that it worked (see *Nature* **394**, 514; 1998). Hundreds of patients may have suffered through not getting better treatment.

‘Breakthroughs’ are regularly announced by the most prestigious Italian newspapers, magazines and television channels. And many times, after yet another explosive report of a remarkable discovery, I have had to disillusion my restaurateur friends about the prospect of eternal youth and similar preposterous claims.

The tabloidesque Italian media have provided us with many entertaining

evenings in New York City. Yet I am left with a bad taste in my mouth despite the excellent dinners I have enjoyed, as I am afraid that the Antonios, Giovannis and Marcellos, unable to distinguish which scientific information is correct and which isn’t, will end up regarding all of it with scepticism. Is it really worth attracting a larger audience on one occasion, if the next time your credibility will be compromised? This is clearly a question for the media as

well as for those scientists who like to promote their discoveries.

Just a suggestion: if you are a scientist in Manhattan, don’t make it evident in certain Italian restaurants. They might not take you seriously, and don’t be surprised if they ask you to pay in cash ...

Michele Pagano

Department of Pathology, New York University
School of Medicine, 550 First Ave, New York,
New York 10016, USA

Industry and evaluation

Sir — Despite the statement in your Opinion article “Finding a future for GM crops” (*Nature* **414**, 1; 2001), the Supply Chain Initiative on Modified Agricultural Crops (SCIMAC) is not involved with coordination of farm-scale evaluations of genetically modified (GM) crops in Great Britain, nor has industry been involved in setting the questions or research agenda underlying the evaluations.

The research agenda was developed by the UK Department for Environment, Food and Rural Affairs (DEFRA) and the Scottish Executive, in response to concerns raised by English Nature and the Royal Society for the Protection of Birds (L. G. Firbank *et al.* *Nature* **399**, 727–728; 1999). The research is conducted by contractors (Centre for Ecology and Hydrology, Institute of Arable Crops Research and the Scottish Crop Research Institute), fully funded by the UK government.

Evaluations are overseen by a steering committee of independent scientists. SCIMAC’s role is to provide a pool of farms from which we scientists select our sample, and which is subject to approval by the steering committee. SCIMAC holds the contracts with the farmers to grow the crops, and is obliged to ensure compliance with the conditions for GM-plant release. SCIMAC has a limited role in providing advice to the farmers on use of herbicides on GM crops. SCIMAC’s role is carefully limited, largely to have a clear division of responsibilities, but also because all parties involved agree that clear boundaries help to avoid any apparent conflict of interest, as your Opinion article stated.

The report by the Agriculture and Environment Biotechnology Commission to which you refer was supportive of the evaluations. Perhaps we scientists were not attacked by the commission precisely because the members appreciated the precautions that have been made to ensure the impartiality of the research.

These precautions are listed in the documentation (see www.defra.gov.uk/environment/fse/index.htm).

Les Firbank

Farm-Scale Evaluations of GM Herbicide-tolerant Crops, Centre for Ecology and Hydrology, Merlewood Research Station, Grange-over-Sands, LA11 6JU, UK

We stand by our account — Editor, *Nature*.

Dogs won more fame than female colleagues

Sir — Your informative News Feature “Eyes on the prize” (*Nature* **413**, 560–564; 2001) contains a photo of the Russian physiologist Ivan Pavlov, winner of the 1904 Nobel prize, with male colleagues and one of the dogs he used to discover the existence of the conditioned reflex.

Some years ago, I came across the original photograph, which also shows two women from Pavlov’s team. One may have been Dr Maria Kapitonovna Petrova and the other Dr Maria Nikolayevna Yerofeyeva. Perhaps your readers may have definite information about their identity?

The existence of these women left on the cutting-room floor has been, like so many others, excised from history. Women scientists deserve better treatment today.

Caroline L. Herzenberg

1700 E. 56th Street 2707, Chicago, Illinois
60637-5092, USA



Full picture: female physiologists, thought to be M. K. Petrova and (far right) M. N. Yerofeyeva.

The poetic mystery of dark matter

The life and poetry of astronomer Rebecca Elson.

A Responsibility to Awe

by Rebecca Elson

Carcanet: 2001. 159 pp.

£6.95 (pbk)

Ingrid Fiske

How can science relate to poetry? Throughout his career, the late Miroslav Holub, Czech clinical pathologist and celebrated poet, probed the relation between the two disciplines, resolutely refusing to acknowledge a difference between the scientific and artistic mind. But it is unusual to find a professional scientist who is also an accomplished poet. Holub himself was one, as was the South African microbiologist Douglas Livingstone. The species is rare.

Rebecca Elson was an astronomer and a poet. She transformed the 'dark matter' of her scientific study into poetic mystery, even into the mystery of her own impending death. Her research centred on globular star clusters, and it was while working at Princeton University on the first data from the Hubble Space Telescope that she began writing poems seriously. In *A Responsibility to Awe* she attempts to understand the relation of her scientific vocation to her poetic one. Experimental and poignant, the poems seek through the pressure of imagery and rhythm to penetrate "the untranslatable dark".

The book has three parts. The first is a selection of Elson's poems; then follows a section drawing on her notebooks; and the volume concludes with an essay Elson originally wrote as a contribution to a collection of autobiographical essays by women alumnae of the Bunting Institute at Harvard University's Radcliffe College.

Elson began writing poems in the back of her geologist father's fieldwork camper. He was the key influence on her scientific education, and in "Poem for My Father" she pays homage to her inheritance: "You honouring all my questions/With your own."

Later, she considered forsaking astronomy for a writing career, but remained faithful to her scientific vocation. Her reason: "I also loved the sense of exploring not an inner, but an outer world, that was really there, in some objective sense". This relation between the inner and outer world preoccupies her poetry.

In the notebook extracts, Elson illuminates contemporary cosmology, and meditates on dark matter and the origins of life. She exhorts: "Don't ask the questions you've been taught by science/Ask it everything you ever wanted." And she warns those "with the power of science" not "to condemn poetry, the spirit/Because curiosity, after all, is also of the spirit", and



Elson's poetry seeks to penetrate the 'untranslatable dark'.

asserts that "explanation is not understanding". Understanding comes through vigilant attention to the sensual world, through fidelity to the spirit and to the way our personal world interacts with the explanatory world of science.

As the notebooks develop, Elson touches increasingly on the fragility of life, and on the nature of ignorance and loneliness "on the edge of the unanswerable". When she discovers she has cancer she enjoins herself to "accept the grace of being a memory". But sometimes her stoicism falters. In one of the poems in the notebooks she talks of life as being a "furrowed field"—looking at freshly turned earth reminds her of "fresh turned graves". Myths of endurance and heroism compel her; many fragments from the notebooks circle around beauty, continuity and death.

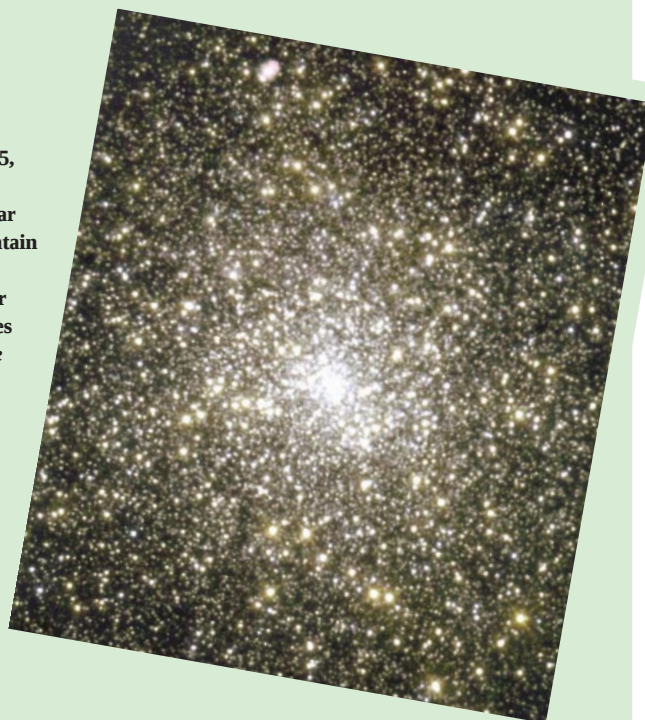
Elson's response to her mortality is uniquely registered through her knowledge of science. "In me now," she writes, "Are traces of the Madagascar periwinkle/Mustard gas/And mutant genes/And things made inside mice." This presages the startling poem "OncoMouse, Kitchen Mouse", with its offer to the mouse "whose cousins gave/their many lives for me" to move into her kitchen, "Make it yours./ Eat all my crumbs". And in the moving fragment "Transumanza" she asks: "Is there any language, logic/Any algebra where death is not/The tragedy it seems/A geometry that makes it look/Alright to die."

Many of Elson's poems found their source in the daily jottings she made, out of which patterns and concentrations of meaning were shaped. In the completed poems the more ragged representations of the notebooks take on a formal, though often still experimental, structure. Awe at the mechanics of the world, meditations on the relation of scientific method to poetic language, and a microscopic focus on the quotidian become recomposed. The science of discovery becomes subject to another discipline: that of poetic technical resources.

"We Astronomers", the first poem of the volume, reads as Elson's manifesto, adumbrating the relation between observation, investigation and awe. A cluster of poems deal with love, the body and the family,

A nebulous secret revealed

The globular star cluster M15, lying in the constellation Pegasus, was the first globular cluster that was found to contain a planetary nebula (seen towards the upper left corner of the field). The image comes from *Cosmic Butterflies: The Colorful Mysteries of Planetary Nebulae* by Sun Kwok (Cambridge University Press, £20, \$29.95), which describes what we know of planetary nebulae—the end stage of a star's evolution.



dwelling particularly on ancestry, cultural differences and evolution. Other poems invoke children, or their questions about the nature of the world. In "Girl with a Balloon", for example, a child carries "A little bit of pure Big Bang/Bobbing at the end of her string". Here the lyric moves characteristically from the social and scientific (the universe in the poem is an industrial warehouse) to the pure, the suspended — a kind of metaphor for poetic activity.

Not unexpectedly, many of the poems refer to stars — but seldom in the unfocused, generic way most poets invoke them. "Aberration" is written about the damaged Hubble telescope. "Constellations"; "Observing" and "Dark Matter" refer to actual scientific discoveries or practices. The last is one of a series of delicate, haiku-like observations of a single natural form or process. Meticulous attention is transformed into wonder. The haunting "Let There Always Be Light" once again reconciles her own poetic enterprise with her scientific vocation: "For this we go out dark nights, searching/For the dimmest stars,/For signs of unseen things." Only occasionally does abstraction, or too extensive use of simile, ground a poem.

Two final poems carry a special poignancy in the light of the poet's death at 39. The striking "These Two Candles, Saint Pantelehm" employs symbolic motifs: hunters drinking in a taverna, a surrealist accident in which the poet sees herself dying, an olive tree "in the hour of a siesta", and an altar where the poet and her companion light candles to Saint Pantelehm. In its stark anticipation of death it offers no comfort. But the final poem, "Antidotes to Fear of Death", does offer an antidote, to both writer and reader — the traditional poetic consolation of continuity in nature, here given a harder edge by the poet's scientific knowledge. Elson reveals that: "Sometimes as an antidote/To fear of death, I eat the stars," and, "Sometimes, instead, I stir myself/Into a universe still young/Still warm as blood."

Softened by an echo from the nineteenth-century poet Gerard Manley Hopkins, the poem ends with quiet, organic trust:

*And sometimes it's enough
To lie down here on earth
Beside our long ancestral bones:*

*To walk across the cobble fields
Of our discarded skulls,
Each like a treasure, like a chrysalis,
Thinking: whatever left these husks
Flew off on bright wings.*

The poems have the same bright wings. ■
Ingrid Fiske is at the Centre for Extra-Mural Studies, University of Cape Town, Rondebosch 7701, South Africa.

The sorry story of drug prohibition

The Pursuit of Oblivion: A Global History of Narcotics 1500–2000

by Richard Davenport-Hines

Weidenfeld & Nicholson: 2001. 481 pp. £20

Leslie Iversen

Our inability to control the illegal use of narcotic drugs, and the ugly criminal underworld that has developed to supply them, represents one of the great unsolved social dilemmas of our age. In *The Pursuit of Oblivion* Richard Davenport-Hines, a professional historian, gives a masterly and readable account of the early history of narcotic drugs, how they came to be widely used in the West in the eighteenth and nineteenth centuries, and how our current draconian drugs-prohibition policy emerged during the twentieth century.

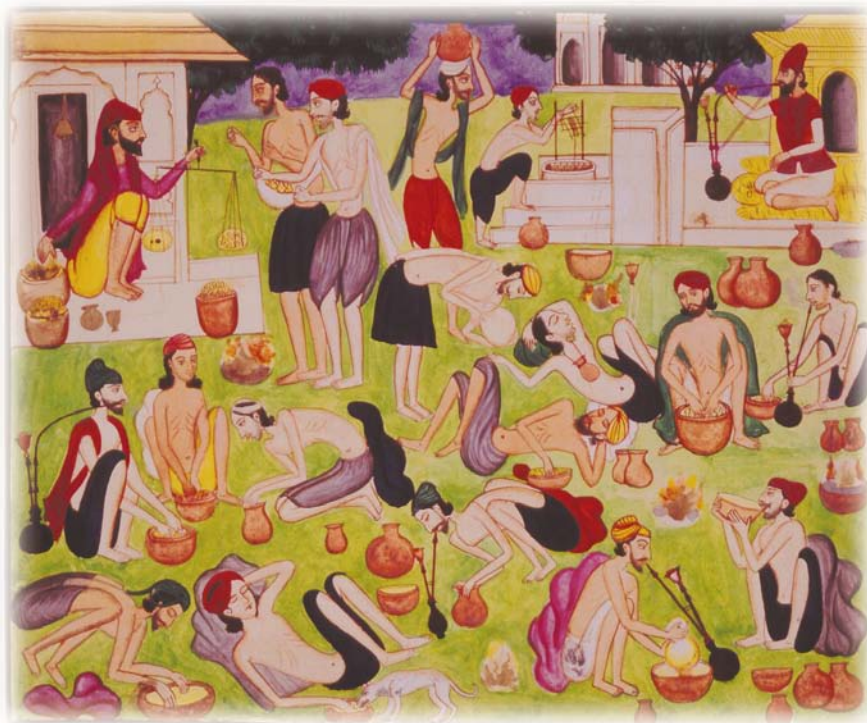
The medical and recreational use of opium and cannabis are part of the ancient cultures of India, Asia and the Middle East, but are relatively recent introductions in the West, where a large trade in opium for non-medical uses developed during the eighteenth and nineteenth centuries. Opiate use was widespread in nineteenth-century Britain, which imported 50,000 kilograms of raw opium in 1827, mainly from Turkey. A large range of opium-based preparations were available cheaply over the counter: laudanum for the gentry, and opium cordials, soothing syrups, elixirs and patent

medicines for all — including some preparations specially designed to keep children quiet during long journeys or while their mothers went to work in the mill.

Several famous orators, including the lawyer Thomas Erskine, the politician William Gladstone and the reformer William Wilberforce, were in the habit of chewing an opium pellet before giving a major speech. Many other eminent people in the arts and politics, including Samuel Taylor Coleridge, Elizabeth Barrett, Lord Clive of India and King George IV were habitual opium users. The problem of opiate addiction was real, but most users did not become dependent. Thomas De Quincey's 1821 cult book *Confessions of an English Opium-Eater* signalled the first alarm about the dangers of abuse, and measures to restrict the use of opiates began to be brought in around the end of the nineteenth century on both sides of the Atlantic.

Meanwhile, cocaine had become available as a pure chemical, and was at first widely adopted as a medicine — it was used by Sigmund Freud both on himself and on his patients. The German drug company Merck sold more than 80,000 kilograms of cocaine in 1885. And new synthetic agents — including ether, chloral hydrate and nitrous oxide — were first used for pleasure and only later as anaesthetics.

The twentieth century saw the rise of an era of strict prohibition — often driven by individuals with strong religious beliefs who saw drug-taking for pleasure as a morally corrupting vice that should be punished. In Britain, Malcolm Delevingne at the Home Office ensured that controlling narcotics



Stimulating societies: narcotics were part of ancient Eastern cultures long before they reached the West.

Going organic

Question: when is a house not a house?
Answer: when it's designed organically.

For then it is more than a house, says David Pearson in his book *New Organic Architecture: The Breaking Wave* (University of California Press, \$60, £40 (hbk); \$35, £24.95 (pbk)) — it is “an organism, an indivisible whole, and humans are seen as part of nature, not above her”. The ‘organism’ shown here overlooks the ocean, and its flowing motifs are meant to reflect the sea’s proximity.



was a police matter and not a health issue. In the United States, Richmond Hobson, Bishop Brent and Harry Anslinger — who headed the Federal Bureau of Narcotics for more than 30 years — were among the prime movers. They refused to see drug abuse as a medical problem that afflicted particularly the socially deprived poor. This policy culminated in the era of Richard Nixon and Ronald Reagan with the ‘war on drugs’.

Davenport-Hines gives a scathing account of the failures of the US drug policy. It generated a new capitalist industry of illicit supply and was responsible for virtually destroying the social fabric of Colombia and other Latin American countries. Among the greatest blunders was the policy adopted by the CIA in Afghanistan, where the rebels fighting the Soviet Union were encouraged to finance their war by sales of opium. By 1980, two-thirds of all supplies of heroin in the United States came from Afghanistan. The policy of banning marijuana as a dangerous narcotic also generated disbelief in a generation of young people who no longer took any government advice on drugs seriously.

Things were only slightly better in Europe, although the British and Dutch were initially sceptical about policies banning opium use in their Far Eastern colonies, where such use was centuries old. At home, freedom of manoeuvre was increasingly restricted by international agreements on drug prohibition through the League of Nations and later the United Nations. But Britain had always viewed the medical model of addiction sympathetically, and heroin addicts were registered and given maintenance treatment until the 1970s. The Dutch also broke away to experiment with the decriminalization of cannabis and its supply through registered ‘coffee shops’. Thankfully, drugs policy is now

being revisited in Britain.

What the sorry history of drug prohibition shows is that this policy doesn’t work. The consumption of cannabis, heroin, cocaine and the new designer drugs ecstasy and γ -hydroxybutyrate (GHB) continues to grow. To treat drug dependence, we need urgently to understand more about the underlying changes in brain chemistry — rather than punishing people. ■

Leslie Iversen is in the Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK.

From clay to computer screen

A History of Writing

by Steven Roger Fischer

Reaktion Books: 2001. 352 pp. £19.95, \$29.95

Maurice Pope

Writing is a sophisticated accomplishment. Was it therefore the product of intelligent design as the eighteenth-century churchman William Paley famously argued, or was it gradually cobbled together by Richard Dawkins’ Blind Watchmaker? Are the hundreds of different scripts still in use each a special creation, or are they branches of a single tree? Steven Fischer stands firmly on the side of interdependence, accident and chains of minor improvement. *A History of Writing* traces the geographic spread of literacy from the Middle East to the world at large, including China and Central America, and its conceptual development from tokens and pictograms to syllabaries and alphabets. The book considers the diverse social and

political implications of literacy, and concludes with some interesting speculation on the possible future of writing in an age of computer icons and voice recognition. All readers are likely to learn something new or to encounter a question they had never thought about before.

Indeed, a number of Fischer’s facts and ideas are new in their own right. The most interesting of these is the attractive theory — originated in 1966 by Pierre Amiet of the Louvre and most actively promoted by Denise Schmandt-Besserat at the University of Texas at Austin — that writing began with accountancy symbols: first, clay tokens, then clay tokens sealed in clay envelopes, then with a count of the contents scratched on the outside as a kind of ‘bill of lading’ to guarantee against fraud. The theory is new because the tokens, small lumps of clay, are not impressive artefacts in their own right and were overlooked by earlier archaeologists.

It is often assumed that what wins must be best. But this is questionable. Arabic numerals may have found favour in the West not because they made calculation easier but because they saved space. The alphabet is not inevitably more efficient than a syllabary; it depends on the language. And our present Latin alphabet, even though it now bids fair to conquer the world, is not the best possible one. It has only 26 letters, whereas most languages have around 40 significant speech-sounds or phonemes; its duplication of upper- and lower-case lettering is quite unnecessary, and the letter shapes themselves — however explicable from a historical point of view — are quite irrational.

Ugaritic, a writing system used on the Syrian coast in the fifteenth century BC, was more intelligent: it had 30 signs constructed on a minimalist scheme, and those that were quickest to draw were allocated (or so it would seem) to the most frequent sounds. And in the fifteenth century AD, Korea’s Hankul script was a syllabary created on advanced linguistic principles, with its letters designed to show the position taken by the vocal organs in articulating the sounds — a unique feature. Yet even though it was promulgated by royal authority — by Sejong, fourth king of the Yi dynasty — it was despised by traditional scholars and its use only became general after the Second World War, when changes in spoken language had rendered it a less perfect instrument. Indeed, as Fischer shows, social prestige and political power have always outweighed intrinsic merit in deciding which scripts become dominant. As for reforming the highly inconsistent spelling habits of present-day English, not only have proposed reforms never succeeded in practice, but they may even be undesirable in theory. Fischer can quote authorities as diverse as Noam Chomski and A. P. Herbert in favour of leaving the apparent chaos alone.

The book has faults. There are signs of

haste — lists ending lamely with an “etc.” or “and so forth”. Dating the Indian Kharosthi script to “the first few centuries BC” is careless writing when what he means is the last few. It is not always clear when Fischer is writing on his own authority and when he is following a secondary source. He can even send us to a general book on the alphabet written 60 years ago as his support for saying that the Arabic and Hebrew scripts are both historically tied to a particular religion. This is odd rather than sinful. But to tell us something that is not generally accepted — such as that Cyprus is the place of origin of the Greek alphabet — without providing evidence, is less excusable. It is worse still when Fischer states that the undeciphered pre-Mycenaean Cretan scripts (including the notorious Phaistos Disk) have been deciphered as Greek. The ‘decipherment’, however, is Fischer’s own and hardly anyone else accepts it. The claim is one that should not have been made in a general book where it is likely to mislead non-experts.

But unwavering orthodoxy would be dull. What is wonderful is to see a subject that embraces so much of human civilization handled with the wide knowledge and breadth of vision that it deserves. ■

Maurice Pope is at 26 Lathbury Road, Oxford OX2 7AU, UK.

New in paperback by the same author A History of Language

by Steven Roger Fischer
Reaktion Books, £9.95, \$17.95

Golem schmolem

Kiln People

by David Brin

Tor: 2002. 336 pp. \$25.95

Henry Gee

According to Jewish folklore, a sixteenth-century rabbi of Bohemia called Yehuda Loew devised an antidote to anti-Semitic attacks: a vengeful automaton made of clay, animated with a scrap of parchment bearing one of the ineffable names of God. This automaton was called the golem. The story of Rabbi Loew, however, is not unique. The golem is mentioned in the Talmud, that sprawling compendium of ancient Jewish scholarship — in which, among other matters, scholars debated whether a golem was human enough to make a congregation quorate.

The Talmud, then, was arguably the first forum to debate the legal and social status of artificial people. The Western canon caught up in 1818, with the publication of Mary Shelley’s novel *Frankenstein*. Like the passages in the Talmud that talk of the golem, *Frankenstein* is an extended debate on whether the feelings and motivations of artificial people can be judged as equivalent to those of humans.

The modern concept of automata, like the legend of Rabbi Loew, was a thoroughly

Bohemian rhapsody. In 1921, the Czech dramatist Karel Capek staged his play *RUR* (‘Rossum’s Universal Robots’). In Czech, the word *robota* means work, but with overtones of the obligations of serfdom. Capek was the first to use ‘robot’ to mean an automaton in the guise of a human being, created to do the jobs that real-life humans found dull. Since then, robots have been a staple of science fiction, from Isaac Asimov’s robot stories of the 1950s to the droids of *Star Wars* fame and the replicants of *Blade Runner*, Ridley Scott’s film based on Philip K. Dick’s story *Do Androids Dream of Electric Sheep?*

David Brin’s latest novel brings the golem up to date. *Kiln People* is a detective story set in a world in which making golems is as easy as making toast. Universal Kilns, a vast conglomerate, supplies citizens of the future with inert clay mannequins which, when animated with the imprint of the customer’s ‘soul wave’ and baked in a kiln, produce replicas of the user that can be despatched on all sorts of errands, from doing the shopping to industrial espionage and even armed combat. After a light breakfast, you can imprint a few golems, send them off to work, and spend the rest of the day playing golf. At the end of the day, your golems (Brin calls them ‘dittos’) will return home, download their experiences into your brain, and dissolve into a recyclable, clayey sludge.

Kiln People centres around the exploits of gumshoe Albert Morris, who specializes in tracking down counterfeiters who churn out cheap knock-offs of ‘pleasure’ dittos imprinted with the soul waves of celebrity courtesans. Morris is a twenty-second-century Sam Spade, and at one level, *Kiln People* is a stylish *noir* thriller that follows all the conventions of the genre. Morris is the brilliant-yet-flawed loner amid a cast of stereotypes: there are wise-cracking villains, gun-toting mobsters, enigmatic and beautiful ladies with dark secrets and super-wealthy sugar-daddies.

So far, so *LA Confidential*. The twist is ‘dittotech’, which allows Morris — and everyone else — to be in several places at once, packing more incident into a day than most people could handle in a month. The plot is serpentine, to say the least, and the reader has to be able to follow several parallel plot strands as various copies of Morris — and Morris himself — investigate a seemingly endless brew of intrigue, double-cross and treble-bluff. At times the story threatens to collapse into chaos, like a golem at eventide — but it is saved by a use of language as artful as that of Chandler or Hammett. (“Me, I was one of a vanishing breed — the employed,” says Morris. “Why stay in school when you have a marketable skill? You never know when it’ll become obsolete.”) Everything pulls together at the last minute in a suitably apocalyptic ending involving the archetypal mad scientist — but, hey! I’m telling you the plot.



On the jungle beat

A day gecko (*Phelsuma madagascariensis*), below, and a cercopithecus (*Cercopithecus mona*). From *Secret Jungle* by Nicole Viloteau (Flammarion, \$40).



The noir stylings don't disguise *Kiln People's* homage to the postwar 'golden age' of American SF. Many famous SF stories imagine otherwise familiar surroundings coming to terms with a single, world-changing innovation. In *Kiln People* it's dittotech, in the same way that in Alfred Bester's *The Stars My Destination* (Millennium Books, 1999) it is the ability to flip yourself from place to place at will. Bester's anti-hero, Gully Foyle, happens to be a master of quantum jumping, in the same way that Albert Morris has a soul-wave ideally suited to turning out reliable dittos — a kink on which the book turns.

Why is the golem — and the robot, its latter-day incarnation — such an enduring device in literature and legend? The easy answer is that it is a metaphor for the human condition: the golem stands to mankind as mankind stands to God. The first golem was Adam, a man made from clay and animated with the Divine spark: the Hebrew words for 'man' and 'clay' share the same root. It is a mark of mankind's very humanity that the golem, although realistic, is never quite as good as the real thing, for the preserve of creation rests with God himself.

Another recent and remarkable manifestation is a short story by self-confessed 'occasional' SF author Ted Chiang. His "Seventy-Two Letters" (in *Year's Best SF6*, edited by David G. Hartwell; Eos, 2000) takes place in an alternative Victorian reality in which many alchemical conceits are scientific

reality, and factories generate golems for specific purposes, dictated by the Hebrew acrostics inserted into their mouths — matrices of 72 Hebrew letters, arranged by skilled 'software' programmers. It is no coincidence that the more mystical byways of Jewish lore contain several formulae of God's name, the most complicated of which is based on a 72-letter acrostic. And so the story of the golem comes full circle. ■

Henry Gee is a senior editor at Nature.

A celebration of science

The Science Book
edited by Peter Tallack
Casell: 2001. 512 pp. £30, \$45
Fran Balkwill

"The first truly accessible, lavishly illustrated story of science," heralds the cover of this weighty tome. Beginning at 35,000 BC with the origins of counting, *The Science Book* spans the millennia with achievements and advances in astronomy, biology, chemistry, cosmology, evolution, geology, mathematics, medicine, physics, etc., etc. A total of 250 scientific milestones are each awarded a double-page spread: the left-hand page describes, in about 500 words, a particular discovery and the scientist(s) responsible;

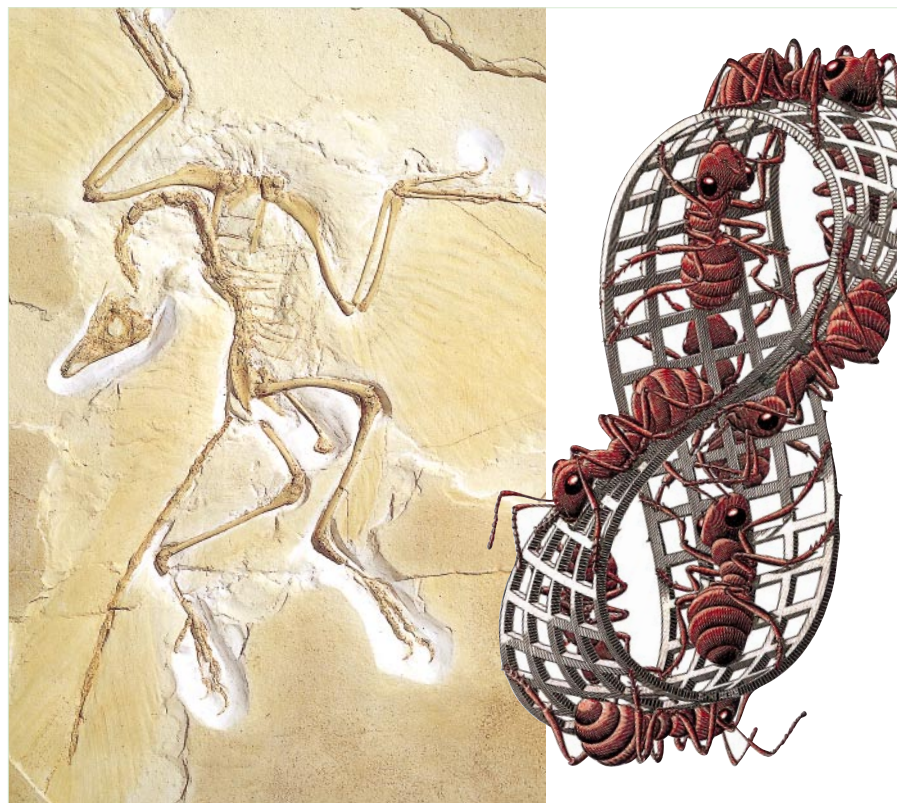
the right comprises a well-selected and often arresting image. There is an enthusiastic preface from Susan Greenfield, a foreword by Simon Singh, and the entries are interspersed with selected writings by an impressive line-up of luminaries, including Peter Atkins, W. Daniel Hillis, Richard Dawkins, Jared Diamond, Richard Leakey, Steven Pinker, Martin Rees and Ian Stewart.

And, yes, the book is accessible. Many prizewinning popular-science books are too difficult for the general reader. But, with less than 500 jargon-free words for each entry, the text of *The Science Book* brings out the salient points in a way that should be clear to most people. The spacious text layout and generous type size certainly draw the reader to each page, and the attention span required for an individual entry is highly suitable for the *Sesame Street* generations.

Science stories are exciting, and science fact is usually stranger and more amazing than science fiction. Yet I found the quality of the individual entries variable. Some were rather pedestrian and, in one case where I know the field, there were worrying inaccuracies. Also, there were omissions. But, as Peter Tallack admits in the introduction, the selection is inevitably subjective, and even science can be a matter of taste. In the field of biomedicine, additional entries on anaesthesia, stem cells, interferon and — especially in view of the recent Nobel awards — the cell cycle would get my vote.

The chronological structure of the book works well, giving an insight into trends and influences across the centuries. This is helped by extensive cross-referencing. But there was no attempt to lead the reader further into the world of science. The index for this celebration of human achievement comprises just one double-page spread. A glossary, bibliography or list of useful Internet sites would all have been helpful. Although this book might well be inspirational for a child or teenager, it may not provide enough detail for that late-night, last-minute, science homework project.

The Science Book is, as it claims, lavishly illustrated, and in most cases the illustrations, which completely cover the right-hand page of each entry, are appropriate, stunning and original. Much can be gained by simply browsing through the pictures, as with other 'coffee-table' books. As I don't have a coffee table, *The Science Book* has been occupying sofa space for the past two weeks. There have been mixed reactions from non-scientist visitors (age range 17–60). Some were encouraging — 'original', 'innovative', 'helps me understand how things were discovered', 'great images', 'useful'. But other comments were not so flattering — 'Reader's Digest book of science', 'a missed opportunity',



The flying and the crawling: *Archaeopteryx*, and red ants demonstrating non-euclidean geometry.

'the design takes me back to the '80s'.

The question that puzzled me was, 'Who would buy this book?' It is not really of encyclopaedic scale or nature; it would surely not be of much use to an A-level science student. Is there a market for coffee-table books on science? I visited my local bookshop, part of a well-known national chain. *The Science Book* was selling well, I was told. Five copies purchased in the first week — good for a non-fiction book of that price. They had just ordered ten more copies for their general stock.

I hope *The Science Book* becomes a best-seller, not just because it celebrates and communicates science, but because there might then be a second edition. It should reduce the size of the images and text, with one entry per page so that 500 scientific milestones/discoveries could be described. The editor should engage as many contemporary scientists as possible in the design, writing and referencing of the book. Then this would be the best popular-science book on planet Earth, and both an inspiration and a celebration for all.

Fran Balkwill is at the ICRF Translational Oncology Laboratory, Barts and The London, Queen Mary's School of Medicine & Dentistry, John Vane Science Centre, Charterhouse Square, London EC1M 6BT, UK.

A clever idea to swallow

This Man's Pill: Reflections on the 50th Birthday of the Pill

by Carl Djerassi

Oxford University Press: 2001. 320 pp.
£12.99, \$22.50

Sexual Chemistry: A History of the Contraceptive Pill

by Lara V. Marks

Yale University Press: 2001. 352 pp.
\$29.95

Michael Gillmer

Fifty years ago, Carl Djerassi and his research assistant Luis Miramontes synthesized norethindrone at the laboratories of Syntex in Mexico City. Norethindrone is the synthetic hormone that has been most widely used over the past half-century in oral contraceptives, and for hormone replacement therapy and treating menstrual disorders. Djerassi's collection of essays is partly historical, partly philosophical and partly autobiographical. Beautifully written, as would be expected of one of the best-known purveyors of the literary genre of 'science in fiction', the essays include Djerassi's views on how 'the pill' has affected the attitudes of both Eastern and Western societies to sexual relationships and reproduction. And they include his

controversial thoughts on the separation of the sex act and human procreation that has occurred as a result of developments in *in vitro* fertilization techniques.

The longest chapter gives Djerassi's account of the development of the pill. Here he repeats a claim first made in *The Politics of Contraception* (1979) that norethindrone, a progestogen, is a contraceptive "pill". But there are several reasons why this does not stand up to close scrutiny. The combined oral contraceptive pill, as it is more correctly known, contains a synthetic oestrogen and a synthetic progestogen. The progestogen used in the initial trials of the pill in the 1950s in Puerto Rico was not norethindrone, but norethynodrel, a product of the Chicago-based company G. D. Searle, for whom Gregory Pincus — rightly acknowledged by Djerassi as the father of the pill — was a scientific adviser.

The inclusion of an oestrogen in the combined oral contraceptive pill was due to serendipity — the original oestrogen component, mestranol, was a contaminant of the synthesis of norethynodrel. The first trials of the pill did not begin until 1956, and it was 1960 before these synthetic hormones were licensed for use in the United States as an oral contraceptive preparation, Enovid. These facts that are fully documented by Lara Marks in *Sexual Chemistry*.

Djerassi's essays describe his views on the control of human fertility and its impact on society, particularly the way in which the development of the pill has affected human behaviour and women's sexual liberation. As he played a major — albeit indirect — role in the changes that have occurred in Western society over the past 40 years, his essays are well worth reading.

Djerassi's autobiography, the curiously titled *The Pill, Pygmy Chimps, and Degas' Horse* (Basic Books, 1992), provides insight into the mind of this enigmatic man. The autobiographical component of this latest book explains his transition from award-winning scientist to acclaimed author, poet and playwright. I can recommend it to both a specialist and a general readership, but suggest that his autobiography is read first as it provides a better insight into Djerassi's early life and influences.

A cursory glance at the main title of Lara Marks' book would suggest a treatise on pheromones and other substances involved in sexual reproduction. But the subtitle clears up that misapprehension. Marks has produced a beautifully written, definitive history of the oral contraceptive pill. Every possible aspect of its development has been considered, ranging from the global population perspective to the impact of the pill on the lives of individual women. Her research

"The Pill is not enough!"



"You need a condom to keep AIDS and STDs away."

You think you know your boyfriend well.
You think you could never get a disease from him.
Think again.
He may not even know he's infected.
Use a latex condom every time.

A condom is for both of you.



Bitter pill: oral contraception has facilitated the spread of HIV.

has been exhaustive, and includes a vast number of interviews with chemists involved in the synthesis of the pill, doctors who conducted the original field trials, and women who took part in these early studies.

Marks has also reviewed all the relevant published literature — both scientific and journalistic — and has scoured the correspondence of the lead players in this fascinating and complex story. Nearly every statement is referenced, and one-third of this 360-page masterpiece is devoted to a detailed bibliography. The book is an invaluable reference source.

Although the subject matter of these two books inevitably covers common ground, the approach taken by the authors could not be more different. Djerassi uses his personal experience and observations as the source of his views — including extracts from his poems, books and plays — to convey his feelings about how the pill has affected women's lives. Marks, on the other hand, has relied very much on the views of others, and ends each of her chapters with a conclusion reached from her extensive research. But both agree that the pill has been a major influence on the lives of women in the twentieth century. And there can be little doubt that it will continue to play an important role in the twenty-first century. These two books in their very different ways provide this insight.

Michael Gillmer is at the Women's Centre, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Sines in terse verse

Coding large numbers in synthetic words made them easier to memorize.

Roddam Narasimha

Here is a bizarre jumble of sounds and/or words — not, as one might think, the rhythm syllables that accompany an Indian dance, but a piece from a renowned Sanskrit work on mathematics and astronomy:

*makhi bhakhi phakhi dhakhi ṇakhi ṇakhi
ṇakhi hasjha skaki kisga ghakhi kighva |
ghlaki kigra hakya dhaki kica
sga s'jha nva kla pta pha cha
kala-ardha-jyāḥ ||*

It is verse 12 from the *Ārya-bhaṭṭīya* written by the great Indian astronomer-mathematician Āryabhata in 499 CE. The sounds are strange not because the language may be unfamiliar; in fact only the last phrase in the verse contains words that you can find in a dictionary of classical Sanskrit.

The verse is effectively a table of sines, and each of the 24 sound-bytes preceding the last phrase is a synthetic word that represents a number, according to a code devised by Āryabhata and explained at the beginning of the book. The code was an ingenious solution to the problem of writing mathematics in terse verse. Verse was the norm in Sanskrit writing, even in science, for it was long before the days of paper and printing, and it had to be terse so that it could be memorized without being an undue burden. Some of these verses are so cryptic that their meaning becomes apparent only from the much longer commentaries — often written in prose — that are the other part of the tradition of Sanskrit writing. Āryabhata was such a powerful and seminal source of tool and thought in Indian astronomy and mathematics that his work attracted a large number of very distinguished commentaries — among the last being one written as late as the nineteenth century by Kodaṇḍa-rāma. It is also possible that the 'terse verse' ensured that the intellectual property did not get lost by falling into whatever was considered 'unauthorized' or 'undeserving' hands.

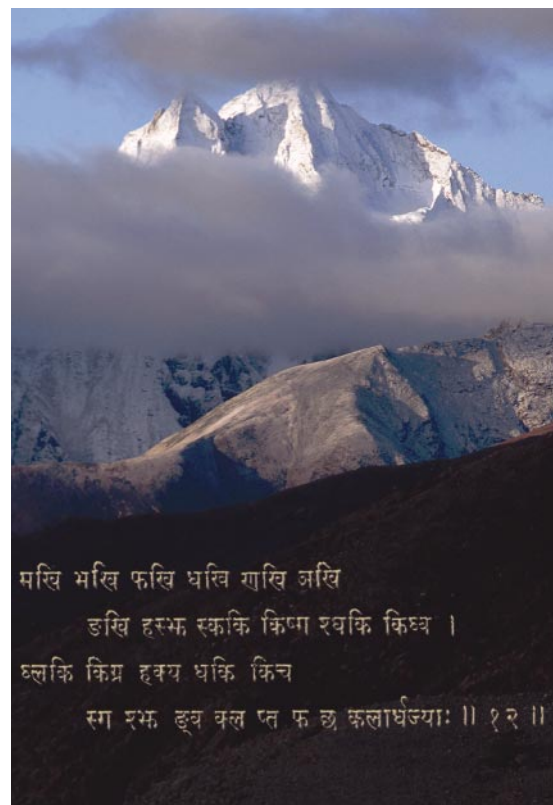
In Āryabhata's system, the 25 'classified' consonants of the Sanskrit alphabet, *k* to *m*, stand for the numbers 1 to 25; the eight unclassified consonants *y* to *h* stand for the numbers 30 to 100 in steps of 10. The place value is indicated by the nine vowels *a* to *au*, (counting long and short as equivalent) progressively from 100^0 to 100^8 in steps of 100 (as the consonants cover 1 to 100, only powers of 100 are needed). A number is denoted by a

simple or compound consonant with a vowel, or a string of such sound-bytes: a compound consonant denotes the sum of the individual ones. Thus $ka = 1$, $ki = 100$ (*i* is the second vowel and stands for 10^2), $gu = 30,000$ (*g* stands for 3, *u* is the third vowel), $gnu = 23 \times 100^2$ (*n* is for 20, $gn = 3 + 20$), and so on. Huge numbers can in this system be represented by short, synthetic words; for example, $khyughr = (4 \times 100^3) + (2 + 30)100^2 = 4,320,000$, which was Āryabhata's postulate for the total number of eastward revolutions of the Sun during an Indian epoch known as *yuga*.

Going back to the sines, I must hasten to clarify that the entries in Āryabhata's verse more precisely provide a set of differences of half-chords for a given radius. As the values are given for 24 divisions of a quadrant, the incremental angle is 3.75 degrees, or 225 minutes; the first difference is *makhi* = $(25 \times 100^0) + (2 \times 100) = 225$, which is of course $\sin 225$ minutes — $\sin 0$, normalized to 225. The actual sines can be obtained by adding up the differences; Āryabhata's values are always correct to three significant figures, very often four. The use of half-chords, rather than the full chord as in Greek practice, was crucial to the emergence of the modern sine (which differs from Āryabhata's in being normalized to the interval 0 to 1). It is not clear why Āryabhata preferred to give differences instead of the sines themselves; perhaps it was because they yielded to a more compact versification, or because they were necessary for extending the tables, or because they helped in interpolation: after all, in 665 CE Brahmagupta offered (in a book called *Khanda-khādhyaka*, literally *The [mathematical] Sweet Eat*) a second-order interpolation formula corresponding to Newton-Stirling; and it is known that classical Indian mathematicians used the 'smooth' variation of differences as a criterion in the final selection of listed values for the function (amusingly, there are copies of the text that have the sines right but the metre is flawed).

Although the use of letters to indicate

Why in India was there the passion to create the shortest phrases for huge numbers?



Sine language: this Sanskrit is a numerical table.

numbers was widespread in the world, most familiarly in Greek and Roman cultures, the Āryabhata system was fundamentally different in two respects. First, the system implicitly recognized the decimal place-value principle, for the rank of the vowel determined an appropriate power of 10. Second, the combination of consonants and vowels made pronounceable synthetic words, which could be integrated into text and metre.

There were other less ingenious syllabic systems in use in India, but without going into them here, we can raise some interesting questions. Why, for example, was there no dread of large numbers in Indian culture? Did familiarity with 10-km-high mountains and kilometres-wide rivers flowing thousands of kilometres across vast plains to huge oceans have something to do with it? And why was there, at the same time, the passion to create the shortest phrases for those huge but friendly numbers? It was said of Indian grammarians that if they could save even half a syllable from one of their rules, they celebrated it like the birth of a son. Sound was holy, and a syllable could hold the infinite: verbal minimalism appears to have gone with numerical opulence.

Roddam Narasimha is director of the National Institute of Advanced Studies, Indian Institute of Science Campus, Bangalore 560 012, India.

FURTHER READING

Shukla, K. S. & Sarma, K. V. *Ārya-bhaṭṭīya of Āryabhata* (Indian National Science Academy, New Delhi, 1976).

Yano, M. *Indian J. Hist. Sci.* **12**, 83–89 (1977).

Blinded by the light

Stan Woosley

If gamma-ray bursts released their energy uniformly in all directions they would be the most powerful events in the Universe. Astronomers are trying to see past the glare to determine the true energy.

Though sometimes touted¹ as the largest explosions since the Big Bang, gamma-ray bursts (GRBs) are looking less energetic than they did just two years ago. Recent observations suggest that the energy required to produce a GRB may not be much more than that of an ordinary supernova, but somehow this energy is focused into a narrow jet moving very close to the speed of light and pointing directly towards Earth. Writing in *Astrophysical Journal Letters*², Frail *et al.* provide evidence, based on measuring the size of the jets, that all GRBs release roughly the same amount of energy.

Gamma-ray bursts are brief flashes of high-energy gamma rays that arrive at Earth from random directions several times a day³. First detected in the late 1960s by US military satellites monitoring clandestine nuclear testing, their nature has long been a mystery, largely because the gamma-ray telescopes used to study them could not give accurate locations. Fortunately, even after a GRB is over (typical durations are 20 seconds) there are glowing embers at optical, X-ray and radio wavelengths that last for weeks or months. Astronomers have to respond quickly and look in the right direction. But since 1997, the rapid detection of these 'afterglows' led to the discovery that many GRBs occur in distant galaxies, usually billions of light years away⁴.

But solving one problem, the location of GRBs, led to another. The energies inferred are enormous, amounting in some cases to the mass of the Sun being converted directly into gamma rays within a few seconds. In the late 1990s model builders struggled to reproduce these sorts of burst energies in computer simulations, and wondered whether the energies really were that large. In particular, would a GRB be just as bright if it were seen from some other angle?

In the past decade, theoretical astrophysicists began suggesting that much less energy would be needed to produce the same brightness if GRBs broadcast their energy in a narrow cone or jet^{5,6}. Most models of GRB formation involve either the cataclysmic collapse of a massive rotating star into a black hole⁷, or a neutron star merging with a black hole⁸. Because the material accreting into the black hole forms a rotating disk, it blocks the outflow of mass, so any material that does escape is forced into two narrow, and oppositely directed, jets. Other astro-

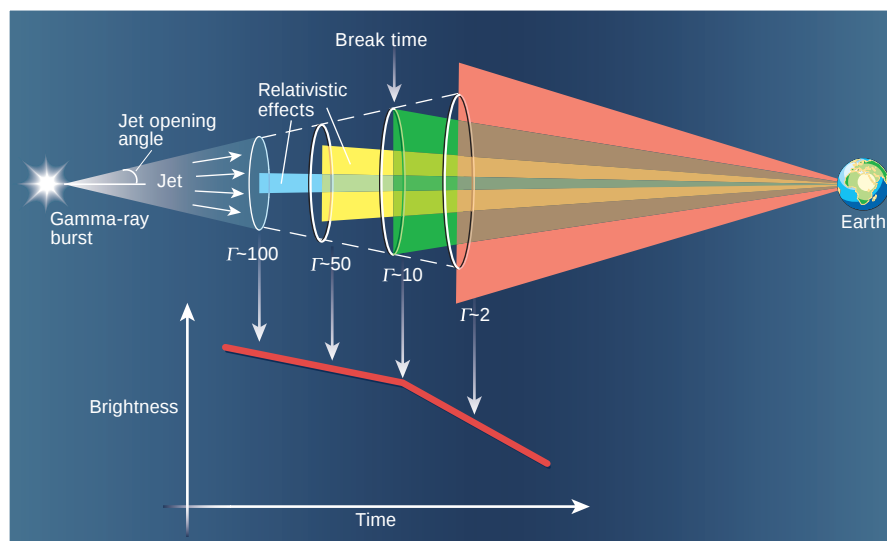


Figure 1 Size and shape of a gamma-ray burst. The burst launches a jet of material that moves at nearly the speed of light, corresponding to a Lorentz factor, Γ , of 200. Because of the effects of special relativity, an observer on Earth can initially only see a tiny fraction of that jet (light blue). Is it a jet or a sphere? It's too early to tell. As time passes and the jet runs into the surrounding material it slows down, and an observer on Earth can see more of the jet (yellow). Eventually, at the so-called break time, the entire jet becomes visible (here, when $\Gamma \sim 10$; green). Beyond this point, no new matter becomes visible and the brightness of the afterglow declines more quickly. When $\Gamma \sim 2$ the jet has slowed enough for all the matter within a much larger area (orange) to become visible. Frail *et al.*² have now analysed the afterglow break times of 17 bursts and calculated the size of their jets. From this they find a typical energy for gamma-ray bursts, which is similar to that of an ordinary supernova.

nomical phenomena thought to be powered by accreting black holes, such as quasars, are also known to produce jets, so why not GRBs? But these arguments were based upon theory and analogy and lacked direct observational evidence.

Frail *et al.*¹ now provide an experimental estimate of the angular size of GRB jets based on observations of their afterglows. Even if we assume that the original emission is spherical, matter released by a GRB is accelerated close to the speed of light and so emits radiation in a narrow beam along its path. The strength of this effect (a consequence of the special theory of relativity) is measured by the Lorentz factor, Γ . A large Γ implies a speed very close to the speed of light, c . So, $\Gamma \sim 200$ means that the matter emitting the radiation is moving at 0.999987 c , and would concentrate its radiation in a narrow cone with an opening angle of 0.28°; matter with $\Gamma \sim 2$ moves at 0.866 c , and would emit radiation in a cone with an angle of 28°.

Producing a GRB⁹ requires Γ to be greater

than 200. So the material emitting the initial gamma rays that we see is only a tiny fraction of the whole; the rest radiates in other narrow cones in different directions. But the space around the burst is not empty, so as the shell of emitting matter moves out it runs into static gas and slows down. Within a week or so it slows enough ($\Gamma \sim 2$) for all the matter moving within a much larger cone to become visible. For many GRBs, the afterglow continues throughout this deceleration phase, so each day we see matter at a larger angle than the day before.

But what if the GRB emission is not a spherical shell, but a jet in the shape of a cone whose base is moving in our direction (Fig. 1)? The cone contains matter that moves at close to the speed of light and produces beamed radiation. Early on we just see a little piece of this cone and so cannot distinguish its emission from that of a sphere. But eventually Γ drops below a certain critical value (depending on the energy of the jet and the density of the surrounding medium)

and we see the 'edge' of the jet. At this point, a plot of brightness against time (light curve) shows a sudden change in slope, growing fainter more quickly because no new emitting material is coming into view (Fig. 1). Such a 'break' in the light curve occurs simultaneously at all wavelengths.

The timing of this break in the afterglow light curve is now known for about 17 GRBs, for which Frail and colleagues² calculate jet opening angles of 2–25°, mostly around 4°. There is a strong relationship between the opening angle and the brightness of the GRB. For example, the brightest GRB known to date, GRB990123, also has one of the smallest opening angles, 2.9°. Frail *et al.* show that the fraction of the sky covered by the jet multiplied by the total energy the burst would have had if it had come from a sphere gives a roughly constant energy value of 5×10^{43} J. This value appears to be typical of most GRBs, and is several thousand times smaller than the energy previously required for GRB990123.

But if a typical burst is visible to only a small fraction of observers, we have to assume that there are far more GRBs than we observe. Frail *et al.* reckon that, for every burst we see, about 500 point in another direction. This means that there are at least 1,500 GRBs every day in the visible Universe, roughly corresponding to the birth of a stellar-mass black hole every minute. This is possible if about 1% of those stars massive enough to die as supernovae end up as GRBs, a reasonable assumption.

This is not quite the whole story — there is still some uncertainty about the efficiency with which the energy of the fast-moving jets is converted into gamma rays¹⁰. Frail *et al.* assume a typical value of 20%, which gives a jet energy of 2.5×10^{44} J. But the efficiency might be expected to vary from event to event, and to depend on the opening angle. They also assume that the density of matter surrounding the GRB is constant, whereas it probably declines with distance. It would also be surprising if the jet were uniform in brightness. Frail *et al.* assume that the opening angle does not vary with time and that Γ does not change with time or opening angle, although models predict that Γ declines from the middle of the jet to its edge and that the angle itself increases with time.

Nonetheless, similar conclusions regarding a standard energy for GRBs have also been reached by others^{11,12}, so it is likely that GRBs, although not standard candles, are almost standard 'bombs'. The jet energy calculated by Frail *et al.* (2.5×10^{44} J) is about twice that of a supernova, but this is just the material moving fast enough to contribute to the afterglow. GRB990123 probably has several times this amount, and even more energy might be released as slower material, especially if the jet has to bore through the remains of its parent star⁷. So the true total

energy of a bright GRB is still probably ten times greater than that of an ordinary supernova, allowing them to keep the title of nature's grandest explosions.

Stan Woosley is in the Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

e-mail: woosley@ucolick.org

1. Kluger, J. *Time* 1 Nov. (1999).
2. Frail, D. *et al.* *Astrophys. J. Lett.* **562**, L55–L58 (2001).
3. Fishman, G. & Meegan, C. *Annu. Rev. Astron. Astrophys.* **33**, 415–458 (1995).
4. Costa, E. *et al.* *Nature* **387**, 783–785 (1997).

5. Meszaros, P. & Rees, M. *Mon. Not. R. Astron. Soc.* **257**, 29–31 (1992).
6. Meszaros, P. & Rees, M. *Astrophys. J.* **397**, 570–575 (1992).
7. MacFadyen, A. & Woosley, S. E. *Astrophys. J.* **524**, 262–289 (1999).
8. Ruffert, M. & Janka, H.-T. *Astron. Astrophys.* **344**, 573–606 (1999).
9. Lithwick, Y. & Sari, R. *Astrophys. J.* **555**, 540–545 (2001).
10. Guetta, D., Spada, M. & Waxman, E.; <http://xxx.lanl.gov/abs/astro-ph/0011170>
11. Freedman, D. L. & Waxman, E. *Astrophys. J.* **547**, 922–928 (2001).
12. Piran, T., Kumar, P., Panaitescu, A. & Piro, L. *Astrophys. J. Lett.* (in the press).

The human genome

Part three in the book of genes

Masahira Hattori and Todd D. Taylor

Working out the draft sequence of the human genome was a landmark achievement. But there's lots more to be done before the finished product is available. The complete sequence of chromosome 20 sets us on the way.

Believe it or not, the Harry Potter stories aren't the only highly anticipated series being published these days. On page 865 of this issue¹, you can find the third instalment in another such series — the book of human genes. This book is being produced by thousands of people around the globe, and began two years ago with the publication of the complete DNA sequence of chromosome 22 by part of the publicly funded International Human Genome Sequencing Consortium². The second instalment³, detailing the sequence of chromosome 21, came in May 2000, and in February this year we had two previews of the whole human genome in draft form — one produced by the publicly funded project⁴ and the other by Celera Genomics⁵.

Chromosome 20 is the third chromosome to be completely sequenced as part of the public project, and is the longest of the three

that have been finished so far — nearly 60 million bases (megabases), some 2% of the human genome. As Deloukas and colleagues¹ describe, they have sequenced roughly 99.5% of the gene-active ('euchromatic') regions of chromosome 20, leaving only four physical gaps spanning about 320 kilobases. (This excludes the centromere — the main constriction along the chromosome, which is important in cell division but is repetitive and nearly impossible to sequence.) The four gaps may contain sequences that are rich in guanine (G) and cytosine (C) bases and are also difficult to tackle with current technologies. The number and total size of gaps are intermediate between those that remain unsequenced on chromosomes 21 and 22, but given the larger size of chromosome 20 it is quite an achievement. The authors also identified ('annotated') 727 genes on chromosome 20, which gives a density of 12 genes

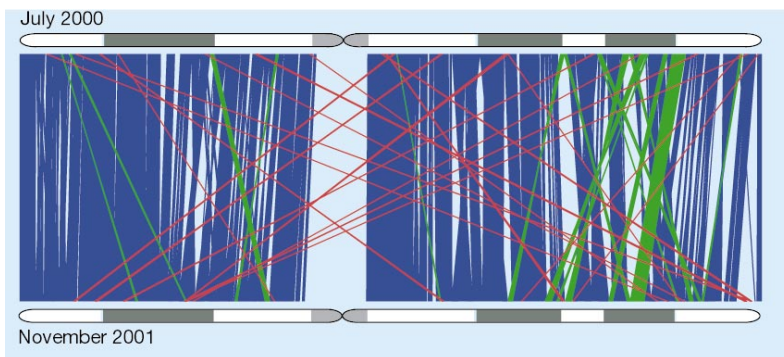


Figure 1 Chromosome 20, from draft to finished quality. The draft data obtained by the publicly funded human genome project⁴ were aligned with the current finished sequence¹. Each line between the two diagrams of the chromosome indicates the same piece of sequence. Colours represent how much the sequence position has shifted between versions (blue, less than 3 megabases; green, 3–10 megabases; red, more than 10 megabases). The ideograms represent the typical banding pattern for chromosome 20, as stained with the dye Giemsa. The light grey region represents the centromere; the white bands are generally rich in G+C bases and dense with genes; the dark grey bands are generally rich in A+T bases and gene-poor. (The figure was prepared by Takehiko Ito, Mitsubishi Research Institute.)

per megabase (compared with the 16.3 genes per megabase on chromosome 22, and 6.7 on chromosome 21).

Deloukas *et al.* analysed the sequence of chromosome 20 with methods similar to those used for the other chromosomes²⁻⁵. But they also introduced new approaches; for example, they compared the sequence with newly released genome data from species such as mice⁶ and puffer fish⁷. Comparing sequences between evolutionarily disparate species reveals genomic sites that have not been able to change much during evolution, because they have essential functions. Such sites are often genes or gene-regulatory elements, and most (even though they are often shorter than 50 bases) are found at similar positions in different genomes. So comparative analyses can be superior to more traditional methods such as computer predictions in allowing unknown genes and regulatory elements to be annotated. And comparative analyses also allow researchers to evaluate the results obtained by these traditional methods. All in all, Deloukas *et al.* suggest that they have annotated as many as 95% of the genes on chromosome 20.

The finished sequence and annotation¹ will be a boon to researchers interested in the diseases that have been linked to chromosome 20. The genes that cause Creutzfeldt-Jakob disease and severe-combined immunodeficiency (an autoimmune disease) were identified and mapped to chromosome 20 several years ago; the genes have since been well characterized.

But many other disorders caused by mutations in single genes or a variety of genes — including type 2 diabetes, obesity, cataracts and eczema — have also been linked to this chromosome. Although the exact genes have not yet been identified, the complete sequence will be invaluable, particularly as researchers have unlimited access to, and are able to freely use, the data. Even while the draft genome sequence was being generated, the discovery of several disease genes linked to chromosome 20 — including those that underlie Hallervorden-Spatz syndrome⁸ (a nerve-degeneration disorder) and Alagille syndrome⁹ (which is characterized by narrowing of the vein leading to the lungs) — was accelerated because the data were freely available. This was one of the goals of the public project from the beginning, and has meant that scientists have also been able to study several forms of cancer associated with chromosome 20. Of course it may take years, but identification of the disease-associated genes should lead, eventually, to a better understanding of the processes that lead to disease, and perhaps to treatments.

Another of the consortium's agreements¹⁰ was that the sequencing of each chromosome would be finished to the same high standards. These standards have been

met for all three completed chromosomes. Although draft data, even at low quality, are useful for initial gene identification and genomic studies, further sequencing and experiments are needed for more precise analyses. It took close to a year to convert the draft data of chromosome 20 to the finished sequence now published. The effort required to polish the last 5 megabases of draft data — and to close 38 of the 42 gaps in the sequence — is comparable to what was required to produce the entire draft. This highlights the fact that 'finishing' has many difficulties, both technically as well as in assessing whether the entire sequence has been covered and the data are of high enough quality.

Finishing includes a tedious process in which incorrect and ambiguous sites in the draft, even those as small as one base, are identified and corrected. Errors such as this have proven recalcitrant to automated sequencing processes, making it clear that the more laborious 'clone-by-clone' approach to finishing is a necessity (as opposed to the whole-genome shotgun approach used by Celera to obtain its draft data⁵). As a good example of this, the highest peak of G+C bases along chromosome 20 occurs at a different position in the finished chromosome 20 sequence¹ and in Celera's draft data⁵. (G+C-rich regions are often difficult to sequence and are of interest because they tend to contain more genes.) There are also major discrepancies between the public consortium's draft⁴ and finished¹ data for chromosome 20 (Fig. 1) — probably due to large duplications in the genome. As

pointed out by Deloukas *et al.*¹ (see Fig. 2 on page 868) and previously^{4,5}, the human genome contains many of these duplications, which might provide the raw material for the emergence of new genes during evolution in higher primates, including humans. Without clone-map information, it is hard to work out where duplicated segments fit in.

As with the Harry Potter series, we already know how long the complete works of the human genome will be — 24 instalments — and we can't wait to get our hands on them all. We have already had a few glimpses of what's to come, but there are many mysteries and secrets yet to be revealed. It's been a long wait between instalments two and three, but there should be a flurry of activity over the next year and a half; we expect the book of human genes to be wrapped up in spring 2003.

Masahira Hattori and Todd D. Taylor are in the Human Genome Research Group, Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-12, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
e-mails: hattori@gsc.riken.go.jp
taylor@gsc.riken.go.jp

1. Deloukas, P. *et al.* *Nature* **414**, 865–871 (2001).
2. Dunham, I. *et al.* *Nature* **402**, 489–495 (1999).
3. Hattori, M. *et al.* *Nature* **405**, 311–319 (2000).
4. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
5. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
6. <http://trace.ensembl.org>
7. <http://www.genoscope.cns.fr/exofish>
8. Zhou, B. *et al.* *Nature Genet.* **28**, 345–349 (2001).
9. Li, L. *et al.* *Nature Genet.* **16**, 243–251 (1997).
10. <http://genome.wustl.edu/gsc/overview/finrules/hgfinrules.html>

Biomechanics

Damper for bad vibrations

R. McNeill Alexander

Some muscle fibres in the legs of horses seem to be evolutionary leftovers with no function. But in fact they may act to damp damaging vibrations generated in the leg as the horse runs.

Horses¹ and camels² have muscles in their legs with tendons more than 600 millimetres long connected to muscle fibres less than 6 millimetres long. Such short muscles can change length only by a few millimetres as the animal moves, and seem unlikely to be of much use to large mammals. The tendons function as passive springs, and it has been assumed that the short muscle fibres are redundant, the remnants of longer fibres that have lost their function over the course of evolution. But Wilson and colleagues argue, on page 895 of this issue³, that these fibres might protect bones and tendons from potentially damaging vibrations.

For mammals and birds with masses greater than a few kilograms, tendon elasticity substantially reduces the energetic cost of

running⁴. In humans, the Achilles' tendon is the principal tendon involved. In animals, other tendons in the lower leg are also important. Each time a foot hits the ground, these tendons are stretched. As the foot leaves the ground again, the tendons recoil elastically. Kinetic and gravitational potential energy are stored as elastic energy in the tendons and returned in the elastic recoil. Similarly, the kinetic energy lost by a bouncing ball when it hits the ground is also stored as elastic energy and returned in the recoil. So running animals effectively bounce along, using less energy than would be needed if their tendons were not elastic.

Compared with other tendons, such as those of the muscles that bend the joints of the leg, the energy-saving tendons have much

smaller cross-sectional areas, relative to the strengths of their muscles⁵. Consequently, in strenuous activities such as running and jumping, much larger stresses act in them than ever occur in more typical tendons. If these energy-saving tendons were thicker, the stresses in them would be lower — but then they would stretch less and store less elastic energy.

In fact, these tendons may stretch so much that the muscle fibres to which they are attached do not need to lengthen or shorten at all as the leg joints bend and re-extend during a step. Experiments on turkeys⁶ and kangaroos⁷ have shown that the bending and extension of the ankle, while the foot is on the ground in a running step, is almost entirely due to tendon stretching and recoil. The muscle fibres remain almost constant in length, doing hardly any work, which suggests that the muscle fibres are superfluous. If so, one would think that the animal could save energy by losing the muscles over the course of evolution: muscles require metabolic energy to develop tension, even when they do not change length and so do no work. In fact, good runners such as antelopes, camels and horses have remarkably short fibres in the muscles whose tendons serve as springs. Yet these muscle fibres have not disappeared.

In their studies of horses' limbs, Wilson *et al.*³ have found a possible explanation. Their experiments show that short muscle fibres can damp the damaging vibrations following the impact of a foot on the ground. When the foot of a running animal hits the ground, the impact sets the leg vibrating; the frequency of the vibrations is relatively high — for example, 30–40 Hz in horses — so many cycles of vibration would occur while the foot was on the ground if there were no damping.

The vibrations might cause damage, because bone and tendon are susceptible to fatigue failure⁸. Fatigue in bones and tendons is the accumulation of damage resulting from repeated application of stresses. Bone fatigue is responsible for the stress fractures suffered by both human athletes and racehorses, and tendon fatigue may explain at least some cases of tendonitis. Wilson *et al.*³ suggest that the very short muscle fibres protect both bones and tendons from fatigue damage by damping out vibrations, but I find the suggestion more convincing for tendon than for bone.

Damage accumulation in tendons is a function of both the total time for which stress is applied, and the number of applications of stress. However, damage accumulation in bones loaded under tension seems to depend only on the total duration of the stress, and not on the number of cycles⁹. Stress fluctuations due to vibrations, superimposed on the overall rise and fall of stress while a foot is on the ground, can be expected to shorten the life of a tendon, but I doubt whether they would have a similar effect on bone.

The findings of Wilson *et al.*³ are original and welcome, and lead one to consider why these tiny muscles might have taken on this role. There is no need to exploit the special properties of muscle to damp vibrations; any viscoelastic material could do the job. For example, if the properties of the tendons were a little different they could be effective as dampers — but then the tendons would not be effective as energy-saving springs. In their elastic recoil, tendons return 93% of the work previously done in stretching them¹⁰, so they have very little damping effect. The 7% of the work that is not returned is dissipated as heat, raising the temperature of the tendons. For galloping horses, the cores of their tendons can reach temperatures high enough to cause damage¹¹, and if tendon properties were modified to make the tendons effective as dampers, the problem of overheating would be worse.

Wilson *et al.*³ have found an important role for a muscle that seemed to be the relic of a structure that had lost its function in

the course of evolution. Their work makes us wonder whether other vestiges (such as the human appendix) are as useless as they seem. ■

R. McNeill Alexander is in the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

e-mail: r.m.alexander@leeds.ac.uk

1. Dimery, N. J., Alexander, R. McN. & Ker, R. F. *J. Zool.* **210**, 415–425 (1986).
2. Alexander, R. McN., Maloiy, G. M. O., Ker, R. F., Jayes, A. S. & Warui, C. N. *J. Zool.* **198**, 293–313 (1982).
3. Wilson, A. M., McGuigan, M. P., Su, A. & van den Bogert, A. J. *Nature* **414**, 895–899 (2001).
4. Alexander, R. McN. *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, 1988).
5. Ker, R. F., Alexander, R. McN. & Bennett, M. B. *J. Zool.* **216**, 309–324 (1988).
6. Roberts, T. J., Marsh, R. L., Weyand, P. G. & Taylor, C. R. *Science* **275**, 1113–1115 (1997).
7. Biewener, A. A., Konieczynski, D. D. & Baudinette, R. V. *J. Exp. Biol.* **201**, 1681–1694 (1998).
8. Ker, R. F. & Zioupos, P. *Comments Theor. Biol.* **4**, 151–181 (1997).
9. Zioupos, P., Currey, J. D. & Casinos, A. J. *Theor. Biol.* **210**, 389–399 (2001).
10. Ker, R. F. *J. Exp. Biol.* **93**, 283–302 (1981).
11. Birch, H. L., Wilson, A. M. & Goodship, A. E. *J. Exp. Biol.* **200**, 1703–1708 (1997).

Meteoritics

Life's sweet beginnings?

Mark A. Sephton

Sugars are common components of organisms on Earth. So their discovery in a meteorite from a lifeless part of the asteroid belt has implications for theories of the origin of life.

There are reasons to believe that some of the building blocks needed to start life on Earth may have come from outer space. Meteorites can be particularly useful vehicles in this regard, especially a type called carbonaceous chondrites, which, as their name suggests, are rich in carbon. Cooper *et al.*¹ have been studying two such carbonaceous chondrites, and on page 879 identify the first extraterrestrial sugar-related compounds found in meteorites.

The Murchison and Murray meteorites studied by Cooper and colleagues (Fig. 1, overleaf), like all carbonaceous chondrites, are derived from ancient asteroids that can be considered as a sort of builder's rubble left over from the construction of the planets. As a result, they have avoided being recycled by planetary processes, and are a valuable record of the state of matter shortly after the Solar System was formed. The diverse types of organic matter present within carbonaceous chondrites constitute evidence that as the Solar System formed it was accompanied by a great deal of organic chemistry. 'Organic' in this sense does not refer to matter created by biological processes, but simply to the chemistry of carbon compounds.

In their paper¹, Cooper *et al.* discuss the significance of the organic matter in meteorites for the origin of life, but in a different

context. In 1961, Oró² first suggested that the early Earth inherited substantial amounts of organic matter from the influx of extraterrestrial organic-rich objects. There has since been a growing acceptance that extraterrestrial organic matter, similar to that found in the Murchison and Murray meteorites, provided the raw materials to help kick-start life on Earth. Until now, though, the organic mixtures in carbonaceous chondrites were missing one obvious ingredient from the recipe of life: sugars.

Collectively, sugars and their related compounds are called polyols. Chemically speaking these are compounds that have a number of hydroxyl (OH) groups attached to their carbon skeleton. Sugars are important biologically because they provide the carbon skeletons for many other molecules. They are also a source of energy for organisms and, when combined to form larger molecules, they can act as food stores and give structural support.

Where did the sugar-related compounds in the meteorites come from? Cooper *et al.* discuss the possibility that they were first formed in interstellar space where there are vast and relatively dense clumps of dust and gas called molecular clouds. Starlight could have irradiated icy mixtures of water, ammonia and carbon monoxide that coated the



1.5 cm

Figure 1 How sweet? Analysis of the Murchison meteorite by Cooper *et al.*¹ shows that this ancient and carbon-rich extraterrestrial object contains sugar-related compounds. If such compounds are more common in meteorites than previously thought, then they could have helped to kick-start life on Earth.

surfaces of small dust particles. The resulting reactions may have generated simple sugar-related compounds or their precursors. Later, a small dense core inside one of these interstellar clouds collapsed to form the solar nebula, a rotating disk of dust and gas which preceded the early Solar System. If simple interstellar organic molecules survived the transformation of the nebula into the Sun and discrete planets, they could easily have been caught up in forming asteroids.

What happened next is less clear. But there is ample evidence for the presence and action of liquid water on the Murchison and Murray meteorites in the form of clays and other hydrated minerals that have been converted from dry starting materials. So it is possible that simple interstellar molecules were stewed in water and condensed to form more complex molecules, perhaps helped along by inorganic catalysts.

However, it is well known that meteorites can become contaminated with terrestrial compounds after their fall to Earth. These contaminants have often given scientists the wrong impression about whether, and how, extraterrestrial organic chemistry takes place. In fact, Degens and Bajor³ reported the presence of sugars in the Murray meteorite as long ago as 1962, but they could not exclude the possibility of terrestrial contamination.

For the sugar-related compounds detected by Cooper *et al.*, reassurance can be derived from the fact that they display many of the features of extraterrestrial compounds. For example, larger sugar-related molecules are less abundant than smaller ones. They also display a wide range of structural diversity, including several different molecular arrangements (isomers) that are rare on the Earth. These characteristics are shared by other compounds found in the meteorites, such as amino acids, and suggest that they were formed by a non-biological reaction mechanism.

A powerful technique used by Cooper *et al.* to determine the origin of the sugars relies on stable isotopes. It is well established that the ratios of carbon ($^{13}\text{C}/^{12}\text{C}$) and hydrogen ($^2\text{H}/^1\text{H}$) isotopes in terrestrial and extraterrestrial molecules can literally be worlds apart. Extraterrestrial organic matter is commonly enriched in the heavier isotopes (^{13}C and ^2H). By making such measurements on the mixture of sugar-related compounds in the Murchison and Murray meteorites, Cooper *et al.* show that most have an extraterrestrial origin. However, these measurements only represent an average of a variety of molecules. Determining which

molecules are indigenous, and which are not, will require compound-specific isotopic measurements⁴. Only then will we be able to further constrain our theories of how, in the early Solar System, the first chemical steps were taken towards sweet life.

Mark A. Sephton is at the Planetary and Space Sciences Research Institute, The Open University, Milton Keynes MK7 6AA, UK.

e-mail: m.a.sephton@open.ac.uk

1. Cooper, G. *et al.* *Nature* **414**, 879–883 (2001).
2. Oró, J. *Nature* **190**, 389–390 (1961).
3. Degens, E. T. & Bajor, M. *Naturwissenschaften* **49**, 605–606 (1962).
4. Sephton, M. A. & Gilmour, I. *Mass Spectrom. Rev.* **20**, 111–120 (2001).

Molecular biology

Specifying transcription

Ian F. G. King and Robert E. Kingston

To switch on the right genes at the right times, our cells rely on many different proteins. Some help to remodel the architecture of the genome, and are remarkably choosy about which other proteins they work with.

To fit inside the nucleus of a cell, DNA must be packaged into a compact structure called chromatin. The building block of chromatin is the nucleosome, which contains 147 base pairs of DNA wrapped around a core made up of four histone proteins. Although essential for nuclear architecture, chromatin structure in general and nucleosome structure in particular create a new problem: they stop many of the proteins that carry out transcription (the copying of DNA into messenger RNA), and DNA replication, recombination and repair, from accessing DNA (Fig. 1). The past decade has seen the discovery of several molecular complexes that use energy derived from the breakdown of adenosine triphosphate (ATP) to unravel chromatin and facilitate these processes¹.

On page 924 of this issue, Lemon and colleagues² look at how one such complex activates transcription. Their results underscore the diversity and specificity of these 'chromatin-remodelling' machines. Remodelling complexes can be divided into four classes, of which the SWI/SNF family is the largest and most diverse, and the SWI/SNF complexes are also the largest in size. They use either the BRG1 protein or the extremely similar hBrm as their central 'engine'³, and contain different sets of associated proteins. Lemon *et al.* find that in a purified *in vitro* transcription system, a SWI/SNF complex called PBAF can strongly stimulate transcription, whereas a complex that differs in just one of the associated subunits cannot.

One of the main challenges to researchers

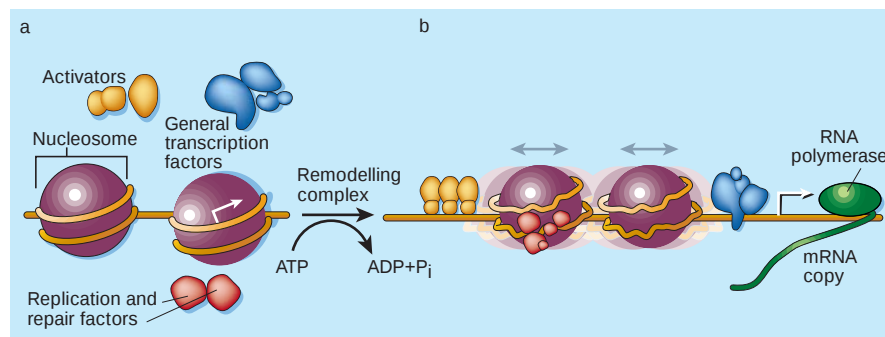


Figure 1 Wrapping up the genome. a, Chromatin structure efficiently packs DNA (orange line) into the nucleus. The basic unit of chromatin is the nucleosome, which consists of DNA wrapped around a core of histone proteins. However, proteins that must access DNA to do essential jobs (such as transcribing genes into RNA copies, replicating the DNA before cell division, and repairing breaks in DNA) compete with nucleosomes for binding sites in DNA. b, Chromatin-remodelling complexes use energy from ATP hydrolysis to make the DNA accessible to these proteins, for example by loosening DNA–nucleosome contacts and sliding nucleosomes to new positions. The white right-angled arrows represent regions in the genome from which the transcription of a gene begins. The double-headed arrows represent potential movement of the nucleosome.

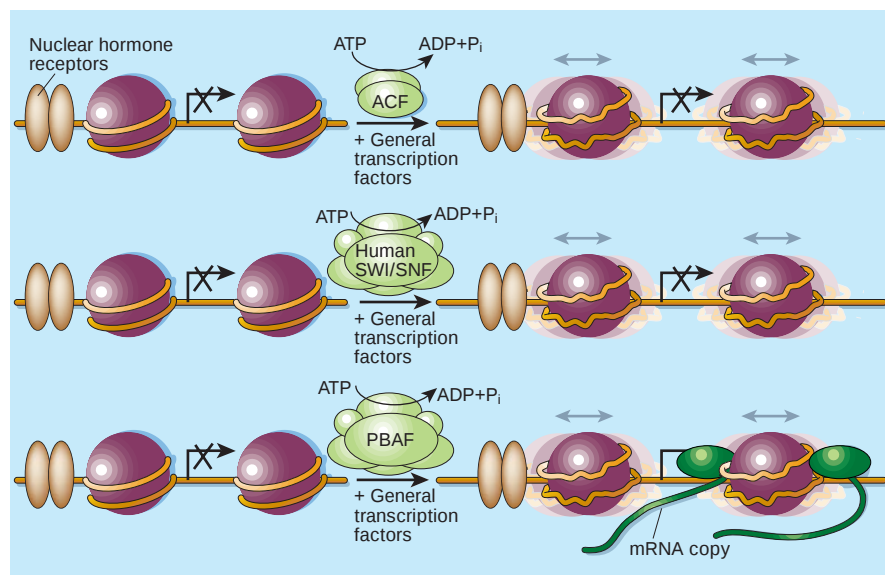


Figure 2 Remodelling complexes and gene transcription. Left, before and right, after remodelling. In the context of a particular class of transcriptional activator protein — nuclear hormone receptors — three different complexes (ACF, human SWI/SNF and PBAF) can remodel DNA that is bundled up into chromatin, as Lemon *et al.* show². But although human SWI/SNF and PBAF differ in only one of their protein subunits, only PBAF can activate transcription in this context.

in this field has been to define the set of protein components necessary to carry out the complex, multistep transcription reaction. The most successful approach to the problem has been to systematically build minimal *in vitro* systems from well-defined and highly purified individual components. Such studies have shown that, when the DNA template is not bundled up into chromatin, several general transcription factors and the enzyme RNA polymerase II (which makes the RNA copy of a piece of DNA) are sufficient to make a specific transcript. Regulation of the amount of transcription is accomplished in part by activator proteins, which require a further set of transcription factors. However, even more factors are required to overcome the blocks to transcription caused by chromatin structure. The factors needed *in vitro* to transcribe a chromatin-bound DNA template have only recently been explored in detail.

Lemon *et al.*² have studied these requirements in the context of a particular class of activator proteins — the nuclear hormone receptors. Reconstituting the machinery required to activate a gene is a considerable technical challenge, and the authors have developed a sophisticated system for doing so. They started with a set of purified proteins that is sufficient to activate transcription on chromatin-free templates, and then investigated which other factors are required for activated transcription of a chromatin-wrapped template. They found that just one further complex was needed — a member of the SWI/SNF family of ATP-dependent chromatin-remodelling complexes known as PBAF (or SWI/SNF-B)⁴.

That an ATP-dependent remodelling

complex is involved in transcription on chromatin is not a surprise. It has been long known that the yeast SWI/SNF complex is involved in transcriptional activation⁵ and can cooperate with nuclear hormone receptors to activate transcription⁶. Further, another remodelling factor — RSF — was previously identified as part of the machinery needed for the transcription of chromatin-wrapped templates in a different minimal *in vitro* system⁷. And SWI/SNF complexes have been shown to enhance transcription from chromatin *in vitro* in less well-defined systems^{8,9}. It is also well known that certain activators can interact with SWI/SNF complexes directly to recruit them to a transcription template¹⁰. For example, the SWI/SNF complex ERC-1 was purified as a factor that interacts with the activator protein EKLF to enhance transcription on chromatin⁸.

Lemon *et al.*'s work² is unique in that they show that only a specific variant of the SWI/SNF family can fully activate transcription in this particular context. Neither a mechanistically different remodelling complex (the ACF complex) nor a complex (human SWI/SNF) that differs from PBAF in only a single subunit was up to the task (Fig. 2). Lemon *et al.* added these different remodelling factors to their reactions, then monitored either chromatin remodelling or the full transcription process. They found that even though all three complexes could remodel the chromatin template, only PBAF could activate transcription. So a specific remodelling complex is required for transcription mediated by nuclear hormone receptors.

It is not clear why the PBAF complex in particular is required. This complex derives



100 YEARS AGO

Messages from Newfoundland announce that Mr. Marconi has succeeded in signalling from England to America by wireless telegraph. Detailed information is not yet available, but it is said that the signals which were received at St. John's, three on Thursday and one on Friday last, though faint were unmistakable, and that Mr. Marconi intends to come immediately to England to increase the power of his transmitters at Poldhu, Cornwall, in order to establish more satisfactory communication across the Atlantic. According to later information the Anglo-American Telegraph Company have given Mr. Marconi notice to remove his instruments from the Colony, as they possess a fifty years' telegraphic monopoly, of which there are still two years to run. This will involve the removal of his experimental station to Nova Scotia or to some other convenient place on the American coast line, and may, perhaps, somewhat delay further experiments. It is to be hoped, however, that we shall before long see a further development of Mr. Marconi's remarkable achievement, upon which if confirmed by subsequent results he cannot be too warmly congratulated.
From *Nature* 19 December 1901.

50 YEARS AGO

In some experimental work which involved the movement of nests within and across the boundaries of male territories in a small nesting colony of red-winged blackbirds (*Agelaius phoeniceus*), R. W. Nero and J. T. Emlen, jun., demonstrated certain characteristics of territorial behaviour which had not previously been described... The males recognized sharp and stable territorial boundaries which could be defined within a few feet and they vigorously defended their territories against the intrusion of alien males, females or fledglings except when the latter were quiet; they also tolerated the introduction of alien nests. The males did not extend their aggressiveness beyond their territory boundaries even in the defence of their mates or nests in neighbouring territories. The females freely followed their nests as they were moved experimentally through the territory of the mate, but assumed a subordinate attitude and followed with difficulty when their nests were transported across territory boundaries; they took no part in the defence of the male territory, but opposed the encroachment of alien females on their nest vicinities.
From *Nature* 22 December 1951.

its name from one of its protein components, BAF180, and perhaps this protein — not contained in the other SWI/SNF complexes studied to date — plays a role in specificity. It is also unclear whether all transcriptional activators display this much selectivity in the additional complexes they use. The viral activator VP16, for instance, can interact with SWI/SNF complexes directly⁹, but a biochemical study using VP16 and a minimal transcription system identified a remodeller (RSF) from a different family as a necessary factor⁷.

An even bigger question is which particular transcription events require ATP-dependent chromatin remodellers such as PBAF. Chromatin structure blocks several of the steps in the highly complex transcription process, and remodellers can help several of them to occur. The need for a specific remodelling complex seen here² may reflect an absolute requirement for a chromatin-remodelling activity in a particular location, time or reaction step to overcome a specific

chromatin barrier. Minimal biochemical systems provide a powerful tool with which to address these detailed mechanistic questions, and work like that of Lemon *et al.* may soon give us the answers. ■

Ian F. G. King and Robert E. Kingston are in the Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, and the Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mail: kingston@frodo.mgh.harvard.edu

1. Vignali, M., Hassan, A. H., Neely, K. E. & Workman, J. L. *Mol. Cell. Biol.* **20**, 1899–1910 (2000).
2. Lemon, B., Inouye, C., King, D. S. & Tjian, R. *Nature* **414**, 924–928 (2001).
3. Phelan, M. L., Sif, S., Narlikar, G. J. & Kingston, R. E. *Mol. Cell* **3**, 247–253 (1999).
4. Xue, Y. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 13015–13020 (2000).
5. Winston, F. & Carlson, M. *Trends Genet.* **8**, 387–391 (1992).
6. Yoshinaga, S. K., Peterson, C. L., Herskowitz, I. & Yamamoto, K. R. *Science* **258**, 1598–1604 (1992).
7. LeRoy, G., Orphanides, G., Lane, W. S. & Reinberg, D. *Science* **282**, 1900–1904 (1998).
8. Armstrong, J. A., Bieker, J. J. & Emerson, B. M. *Cell* **95**, 93–104 (1998).
9. Neely, K. *et al.* *Mol. Cell* **4**, 649–655 (1999).
10. Kadam, S. *et al.* *Genes Dev.* **14**, 2441–2451 (2000).

Geophysics

Deep down at Chicxulub

Jay Melosh

The best-preserved large impact crater on Earth is overlain by a kilometre of sediment. It is possible to look not only through that wrapping but also beyond, at the effects of the impact at the crust–mantle boundary.

About 65 million years ago, an extra-terrestrial body hit what is now the north coast of the Yucatan Peninsula, Central America, to create the Chicxulub crater. This is the youngest and best preserved of the Earth's known 'big three' impact craters. The other two, Vredefort in South Africa (2,020 million years old) and Sudbury in Canada (1,850 million years old), have both undergone severe deformation due to tectonic forces¹. The Chicxulub crater, by contrast, seems nearly pristine, but it is entirely buried beneath nearly a kilometre of sediment. Writing in the *Journal of Geophysical Research*², Gail Christeson and colleagues describe the results of the latest attempts to peer through these sediments and look at the crater structure and beneath it.

Thanks to 50 years of work with various approaches — geological field studies, laboratory-scale simulations, analysis of almost-full-scale nuclear explosions, and numerical modelling — we have a clear idea of how the energy of a high-speed impact creates a smallish crater a few kilometres across. As their size increases, however, craters become more complex and our understanding of how they form declines. Extraterrestrial craters that are 20 km or more in diameter display terraced walls and central peaks. Craters in the 100-km class

develop internal 'peak-rings', and still larger ones are surrounded by inward-facing fault scarps that form a class known as multi-ring basins. Although these morphological transitions are well known from images of craters on the Moon and other planets, we

are almost entirely ignorant of what lies beneath their surfaces.

In contrast, although the surfaces of terrestrial impact craters are often erased by erosion, they sometimes offer a glimpse of what lies beneath. The challenge here is to relate their vanished surface morphology to the exposed deep structure. Unusually, however, Chicxulub was buried shortly after its formation. Its existence was first recognized by geophysical methods³, after which its shocked and melted rocks were sampled in boreholes that just pierced the sedimentary cover⁴.

The work of Christeson *et al.*², a joint endeavour between the University of Texas at Austin and Imperial College, London, involved analysing some 650 km of deep seismic reflection profiles of the Chicxulub basin. The profiles were acquired with data from BIRPS (British Institutions Reflection Profiling Syndicate), along with records from arrays of seismometers set on the ocean floor and from seismic stations on land. The seismic reflection results, previously reported, revealed that Chicxulub is a true multi-ring basin. It possesses an outer fault scarp about 180 km in diameter, a slump terrace zone extending from about 40 to 70 km from the centre and a peak ring about 80 km in diameter (Fig. 1).

The new work uses data from seismic refraction: sound signals that curve on broad arcs through the crust, or are reflected at the crust–mantle boundary, to be received at distant stations. Accurate measurement of travel times permits precise measurement of the speed of sound propagation in the crust and, with modern techniques of 'tomographic inversion', produces a detailed image of the crust. These images are further enhanced by data on the acceleration of

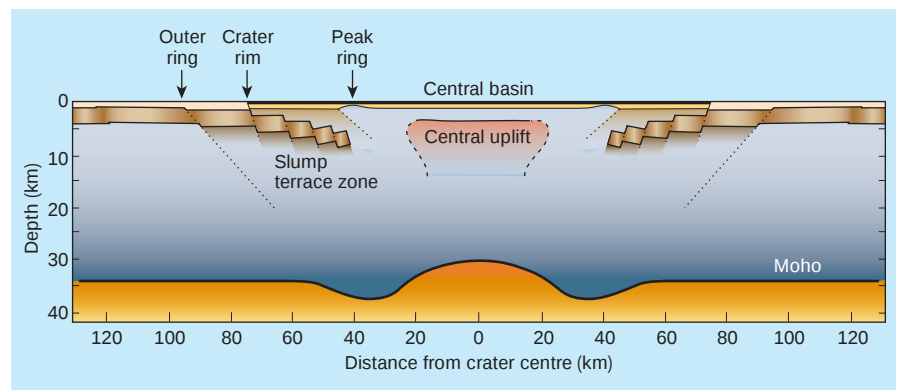


Figure 1 Morphology of the Chicxulub crater. The crater was made some 65 million years ago when an asteroid or comet, 12–15 km in diameter, hit the Earth. The initial, transient crater was bowl-shaped, and 100 km in diameter and 30 km deep, but collapsed within minutes to create the structure seen today. The multi-ring structure consists of an outer fault scarp (outer ring), which possibly formed as mantle or lower crustal material flowed inward to fill the crater cavity; a slump terrace zone extending from the crater ring; and a peak ring, with a radius of 40 km, that encircles an area of central uplift stemming from inrushing rock debris. Although the crust–mantle boundary, the Moho, was first pushed down into the mantle, it rebounded as the crater collapsed. The results of Christeson *et al.*² show that the Moho is now higher directly beneath the crater and lower away from the centre, with the maximum distortion (indicated here in exaggerated form) at about 35 km.

gravity at the surface, which indicate the density of the underlying rock. Knowledge of both sound velocity and density help to distinguish between different rock types.

The new analyses² confirm the previous picture and add some notable new features to it. The first is a low-velocity structure near the crater's surface that is likely to be a disk of frothy impact melt mixed with broken fragments of bedrock. The disk is about 1 km thick and 80–100 km in diameter. It probably does not represent all the melt rocks lying in the crater: undetected denser melt might lie deeper near the centre. The data also reveal a stratigraphic uplift in the crater's centre that varies between 9 and 18 km, in fair agreement with the expected uplift of 18–20 km for a crater of this size.

The most important new data concern the deepest part of the structure, the crust–mantle boundary, or Moho. The refraction data show that the Moho is distorted beneath the crater (Fig. 1). It rises about 1 km above its regional depth of 34 km right under the centre, then sinks about 1.25–1.5 km deeper than average at a distance of about 35 km from the centre. This deepening coincides with the projected trace of the outer ring fault and could reflect inward slip on the fault and flow of the deeper portions of the crater towards the centre. This is the first time that a change in crustal thickness has been found beneath a terrestrial crater.

It might seem surprising that the Moho shows such a tiny perturbation. The crater structure requires that the original, or transient, crater that collapsed to form the present structure was about 100 km in diameter. A transient crater this large has a depth of about 30 km, so shortly after the impact (a few minutes) the crater floor was almost at the depth of the pre-impact Moho. If the crater went nearly all the way through the crust, why is the Moho now nearly where it was before the impact? This apparent conundrum occurs frequently in discussions about cratering, but it has a simple solution.

Although the transient crater was briefly 30 km deep, the material actually excavated from the crust only extended to about 10 km depth. The material beneath the transient crater, about 20 km of crust, was simply pushed down under the crater floor, not excavated. When the crater collapsed this material rebounded, the 10-km-deep excavation cavity was mostly filled in by the lateral inflow of rocky debris, and the Moho resumed nearly its old position. At the same time, layers less deep than the Moho squirted higher beneath the crater floor, finally producing the observed 9–18 km of uplift at these higher levels.

Strange as this train of events might seem, it was first observed in small-scale, explosion-cratering experiments performed in a centrifuge to increase the force of gravity. At first the experimenters believed that the apparent

lack of uplift of layers deep beneath the crater indicated that the crater did not go through a deep transient stage⁵. But later experiments, using photography to follow crater excavation and collapse, showed that the deep layers were indeed first pushed down but then nearly recovered their initial position⁶. Such continuity of layers deep beneath an impact crater (but shallower than the transient crater) is also displayed on seismic profiles of the 3-km-diameter Steinheim crater in Germany⁷.

At Chicxulub a small, but measurable, mantle uplift occurs beneath the crater. This is the only crater on Earth that shows such a feature. If Sudbury or Vredefort once had mantle uplifts, flow in the lower crust and upper mantle has since erased them. Looking out into the Solar System, the largest crater on Venus — Mead, 235 km in diameter — is associated with a negative gravity anomaly that rules out any substantial mantle uplift beneath it⁸. In contrast, the much larger basins on the Moon not only possess mantle uplifts, but some of these uplifts are higher than necessary to compensate for the excavation of material from the overlying crust⁹. Collapse of these monster craters was apparently so vigorous that the lunar mantle was caught up in the rebounding central peak and produced a positive gravity anomaly in the crater centre.

Chicxulub's status as the best-preserved large impact crater on Earth is due to its protection, like that of an Egyptian mummy, by thick wrappings — in this case, of sedimentary material. As with many mummies, its interior is being revealed by modern tomographic reconstructions. In the future we can look for further insights into Chicxulub, not only from improved seismic imaging methods but also from the deep drilling project, organized by the International Continental Scientific Drilling Program, that commenced operations earlier this month at a site named Yaxcopoil-1. With this knowledge we will be better placed to interpret what lies beneath the facial scars on our fellow planets. ■

Jay Melosh is in the Lunar and Planetary Lab-West, University of Arizona, 1040 East Fourth Street, Tucson, Arizona 85721-0077, USA.
e-mail: jmelosh@lpl.arizona.edu

- Grieve, R. & Theriault, A. *Annu. Rev. Earth Planet. Sci.* **28**, 305–338 (2000).
- Christeson, G. L., Nakamura, Y., Buffler, R. T., Morgan, J. & Warner, M. J. *J. Geophys. Res.* **106**, 21751–21769 (2001).
- Penfield, G. T. & Camargo, Z. *Tech. Prog., Abstr. Biogr., Soc. Exploration Geophysicists 51st Annu. Int. Mtg* 37 (1981).
- Hildebrand, A. R. et al. *Geology* **19**, 867–871 (1991).
- Schmidt, R. M. & Holsapple, K. A. *Lunar Planet. Sci. Conf. (Abstr.) XII*, 934–936 (1981).
- Schmidt, R. M. & Housen, K. R. *Int. J. Impact Eng.* **5**, 543–560 (1987).
- Reiff, W. *Guidebook to the Steinheim Basin Impact Crater* (Geologisches Landesamt Baden-Württemberg, Steinheim am Albuch, 1979).
- Banerdt, W. B. et al. *Icarus* **112**, 117–129 (1994).
- Hood, L. L. & Zuber, M. T. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 397–409 (Univ. Arizona Press, Tucson, 2000).

Daedalus

Away with oxygen!

The air-filled pneumatic rubber tyre has revolutionized transport, but is sadly vulnerable to oxygen. When the polymer is oxidized it becomes inelastic, and its surface wears. Vast numbers of worn rubber tyres are thrown out, and new ones bought. Piles of discarded tyres disfigure the landscape, are thrown into the sea as 'artificial reefs' for fish, or are destroyed by burning. Indeed, air defines the whole problem. Tyres run in air, which is 20% oxygen, and are inflated with it. DREADCO chemists now propose to inflate a new tyre with carbon dioxide. This gas is significantly soluble in rubber. It should saturate the polymer, keeping it elastic by displacing the troublesome oxygen. In its slow leak outwards, it will sweep away external oxygen. A cartridge inside will accept liquid carbon dioxide at a pressure of 70 atmospheres, and keep the tyre at its typical 2 atmospheres. It will maintain gas throughput and keep the tyre taut and perfect.

Marketing will be difficult. Modern tyres need little attention (Daedalus greatly admires the advertisement in which the Michelin Man in 'Showboat' dress sings "O! Man Rubber, he just keeps rolling"). And non-leaky butyl rubber has eliminated the tradition of having your tyres pumped up whenever you drive in for more petrol. But smart garages could use carbon dioxide, possibly spiked with oxygen-scavenging additives such as nitric oxide, to inflate tyres. Motorists might reckon that more frequent pumping outweighed the huge cost of changing worn tyres. The ideal gas should save enough rubber to satisfy the market, and solve the terrible problem of endless worn tyres. Furthermore, a really important use for carbon dioxide would do much to counter our fears about releasing this greenhouse gas.

Rubber is also worn on shoes (pun intended). Some training shoes have a gas-blown sole for easier running. A pneumatic shoe pumped up with carbon dioxide could give a very easy 'ride'. We might even get rid of laces, a dreadful invention, by designing shoes that become taut as they are inflated. Now that rubber has replaced leather as a sole material, a wear-proof rubber sole and can-pumped shoe could be attractive. Even better, a small shoe might withstand much more pressure than a big tyre. Narrow carbon dioxide channels could hold the gas in almost liquid state under high pressure, combining long life with a comfortable ride. But such high pressures might make wearers nervous. David Jones

Endothelin-1 synthesis reduced by red wine

Red wines confer extra benefit when it comes to preventing coronary heart disease.

Statistical evidence of reduced coronary heart disease in areas of high wine consumption has led to the widespread belief that wine affords a protective effect^{1,2}. Although moderate drinking of any alcohol helps to reduce the incidence of coronary heart disease^{3,4}, there is no clear evidence that red wine confers an additional benefit⁵. Here we show that red wines strongly inhibit the synthesis of endothelin-1, a vasoactive peptide that is crucial in the development of coronary atherosclerosis⁶. Our findings indicate that components specific to red wine may help to prevent coronary heart disease.

The concept of the 'French paradox' has arisen from reports that deaths from coronary heart disease are much lower in France than in the United Kingdom, despite a comparable dietary intake of saturated fats by these populations^{1,2} — this has been attributed to the higher consumption of alcohol in France, particularly of wine^{1,2}. Mechanisms implicated in this phenomenon include increases in high-density lipoproteins (HDLs) and fibrinolytic activity, and decreased platelet aggregation²⁻⁴, but these changes are modest and can also be caused by ethanol consumption *per se*^{3,4}. Indeed, the very existence of the French paradox has been questioned as it may simply reflect a time lag in dietary cholesterol intake⁴. Identification of a specific property of red wine that accounts for reductions in coronary heart disease could resolve this controversy, as well as providing insight into the health benefits of a Mediterranean diet.

Endothelin-1 (ET-1) was originally described as a highly potent vasoconstrictor peptide⁷, and its overproduction is seen as a key factor in the development of vascular disease and atherosclerosis⁶. Experimental models of atherosclerosis indicate that endothelin antagonists prevent manifestation of the early stages of the disease, such as endothelial dysfunction or fatty-streak formation⁶, and reduce myocardial infarction in established disease⁸. The coronary blood supply of patients with coronary heart disease is also severely perturbed by local ET-1 production⁹. We investigated whether red wine could inhibit the synthesis of ET-1, as this might explain its cardioprotective properties.

We found that polyphenols from red wine made from Cabernet Sauvignon grapes decreased ET-1 synthesis in cultured bovine aortic endothelial cells (BAECs) by suppressing transcription of the ET-1 gene (Fig. 1a). To test whether this property is peculiar to red wine, we prepared ethanol-

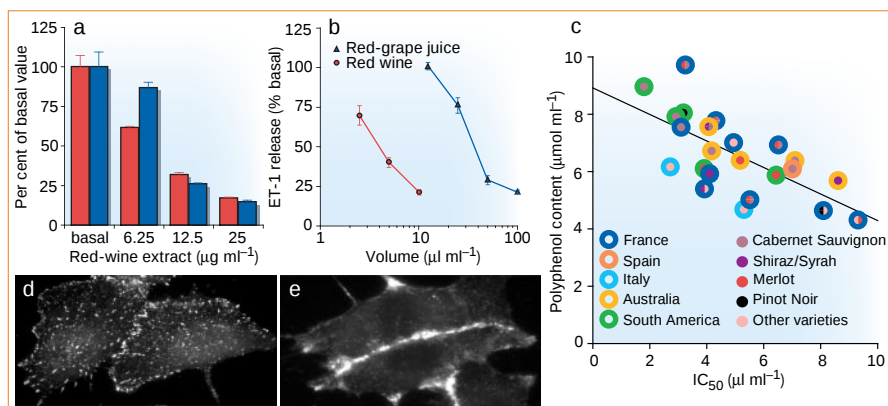


Figure 1 Red wine inhibits endothelin-1 (ET-1) synthesis by bovine aortic endothelial cells and alters the distribution of phosphotyrosine immunofluorescence. **a**, Red wine extract suppresses ET-1 release (red bars) and transcription of the prepro-ET-1 gene (ET-1 reporter gene activity; blue bars). **b**, ET-1 release over 6 h during incubation with extracts of red-grape juice (triangles) and a representative red wine (circles) (plotted as equivalent dilutions to unextracted samples). **c**, Correlation of total polyphenol content of red wine (measured with Folin and Ciocalteu's reagent and expressed as catechin equivalents in $\mu\text{mol ml}^{-1}$) with the concentration of red-wine extract causing a 50% reduction in basal ET-1 synthesis (IC_{50} , expressed as μl -equivalents of unextracted wine per ml culture medium) (see supplementary information for details). **d**, **e**, After 1 h treatment of bovine aortic endothelial cells with either medium or red-wine extract ($50 \mu\text{g ml}^{-1}$), phosphotyrosine (PY)-containing proteins in permeabilized cells were detected as brightly staining regions by using anti-PY monoclonal antibody 4G10 and FITC-labelled goat anti-mouse IgG. **d**, Control cells incubated with medium; **e**, cells after treatment with red-wine extract. Original magnification, $\times 100$.

free extracts from 23 red wines, four white wines, one rosé wine and one red-grape juice (see supplementary information for details). These extracts were tested on BAECs to determine the concentration of each wine that causes a 50% reduction in basal ET-1 synthesis (IC_{50}) (Fig. 1b).

For the red wines, the degree of inhibition of ET-1 synthesis was correlated with the total polyphenol content (Fig. 1c; $r^2 = 0.46$, mean $\text{IC}_{50} = 5.0 \pm 0.4 \mu\text{l ml}^{-1}$). Red-grape juice also inhibits ET-1 synthesis, but is markedly less potent than red wine ($\text{IC}_{50} = 35 \mu\text{l ml}^{-1}$). The white and rosé wines had no effect on ET-1 synthesis ($< 5\%$ inhibition at $100 \mu\text{l ml}^{-1}$). As the rosé wine was from Cabernet Sauvignon grapes, this indicates that the active principle in red wine must derive from red-grape skins or other grape components during the vinification process.

Although polyphenols in red wines are known to have antioxidant properties¹⁰, it is unlikely that this accounts for the effect on ET-1 synthesis. For instance, quercetin ($10 \mu\text{M}$) totally inhibits the oxidation of low-density lipoproteins (LDLs)¹⁰; however, at this concentration neither red-wine polyphenols (quercetin, resveratrol, D,L-catechin, D,L-epicatechin) nor anthocyanins (delphinidin, pelargonidin, cyanidin, peonidin, petunidin, malvidin) affect ET-1 production.

Inhibitors of the cellular tyrosine-kinase family of phosphorylating enzymes that

share structural similarity to red-wine polyphenols also suppress ET-1 synthesis⁶. We therefore investigated whether this action of red wine might be explained by an inhibitory effect on this same family of enzymes by using immunocytochemistry to visualize tyrosine phosphorylation in endothelial cells. Compared with control cells, red-wine extract causes a marked change in cell morphology and a redistribution of phosphotyrosine staining (Fig. 1d, e). These effects on tyrosine phosphorylation are presumably due to modified tyrosine-kinase signalling in endothelial cells.

Red-wine extract is also known to elicit endothelium-dependent vasodilation and lower blood pressure¹¹, which may provide further protection against coronary heart disease. Our findings indicate that remarkably small amounts of red-wine extract can suppress ET-1 synthesis: assuming adequate absorption of the active component, they support assertions that a moderate intake of red wine can prevent coronary heart disease. Characterization of the vascular mechanisms underlying red wine's beneficial effects should help in the design of strategies to prevent atherosclerosis.

Roger Corder, Julie A. Douthwaite, Delphine M. Lees, Noorafza Q. Khan, Ana Carolina Viseu dos Santos, Elizabeth G. Wood, Martin J. Carrier

William Harvey Research Institute, Barts & the London School of Medicine & Dentistry, Queen

Mary University of London, Charterhouse Square,
London, EC1M 6BQ, UK
e-mail: r.corder@qmul.ac.uk

1. St Leger, A. S., Cochrane, A. L. & Moore, F. *Lancet* **1**, 1017–1020 (1979).
2. Renaud, S. & De Lorgeril, M. *Lancet* **i**, 1523–1526 (1992).
3. Doll, R., Peto, R., Hall, E., Wheatley, K. & Gray, R. Br. *Med. J.* **309**, 911–918 (1994).
4. Law, M. & Wald, N. Br. *Med. J.* **318**, 1471–1476 (1999).
5. Goldberg, I. J., Mosca, L., Piano, M.R. & Fisher, E. A. *Circulation* **103**, 472–475 (2001).
6. Corder, R. in *Handbook of Experimental Pharmacology*:

Endothelin and its Inhibitors (ed. Warner, T. D.) **152**, 35–67 (Springer, Berlin, 2001).

7. Yanagisawa, M. et al. *Nature* **332**, 411–415 (1988).
8. Caligiuri, G., Levy, B., Pernow, J., Thorén, P. & Hansson, G. K. *Proc. Natl Acad. Sci. USA* **96**, 6920–6924 (1999).
9. Kinlay, S. et al. *Circulation* **104**, 1114–1118 (2001).
10. Frankel, E. N., Kanner, J., German, J. B., Parks, E. & Kinsella, J. E. *Lancet* **341**, 454–457 (1993).
11. Diebolt, M., Bucher, B. & Andriantsitohaina, R. *Hypertension* **38**, 159–165 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Image processing

Fractals in pixellated video feedback

Video feedback occurs whenever a video camera is directed at a screen displaying the image currently being recorded by the camera. It can be observed in everyday situations, for example at sporting events when a stadium's display screen comes into the camera's view. Here we consider how this simple physical process is affected by the fact that monitors are pixel-based, and show that it can result in stationary fractal patterns such as von-Koch snowflakes and Sierpinski gaskets.

Video feedback is a popular scientific phenomenon^{1,2}, mainly because of its 'beautiful and mesmerizing' images³. It was scientifically investigated in the days of scanline-based cameras and monitors^{4,5} and is often discussed in the context of fractals as an example of a simple feedback process (see ref. 6, for example).

The best known video-feedback phenomenon is perhaps the 'monitor-inside-a-monitor' effect⁷, which occurs when the overall magnification, M , of the camera–monitor combination is less than one ($M < 1$), causing the monitor to display

a tunnel-like, non-fractal pattern consisting of nested images of itself. Video-feedback set-ups can be modified, for example by using multiple monitors⁸ or multiple lenses⁹, in order to produce stationary fractal patterns.

We demonstrate that pixellated, but otherwise unmodified, video feedback with $M > 1$ can lead to fractal patterns (we dub this the 'monitor-outside-a-monitor' effect). Previous experiments with $M > 1$ produced non-stationary complex patterns — for example, rapidly rotating planet-like, fractal-looking structures suspected of being connected to pixels¹⁰. Pixels were also described as acting like the 'cells' of a cellular automaton, a class of abstract machine capable of producing fractal patterns¹¹, and simulations of video feedback on a matrix model — in which the matrix elements acted like square pixels — produced stationary fractal spirals¹².

We predicted that stationary self-similar fractal patterns would be created in pixellated video feedback by analogy with fractal laser modes¹³. These patterns result from iterated magnification and pixellation of the image: the former successively stretches any structure in the image to M , M^2 , M^3 , ... times its original size, whereas the latter continuously adds small-scale structure in

the form of the pixel mask. The result is a pattern consisting of the pixel-mask pattern in various sizes. This hallmark of self-similarity is seen in Fig. 1a: the pattern on the monitor is a large rosette consisting of small rosettes; the small rosettes, each comprising seven bright pixels, are magnified and pixellated images of individual pixels, and the large rosette is a magnified and pixellated image of the central small rosette.

The detailed shape of the stationary pattern depends on the shape and size of the individual pixels, the geometry of the pixel array, the magnification and the position within the pixel array of the centre of magnification. For example, a magnification $M = 2$, combined with a centre of magnification midway between three nearest-neighbour pixels in a hexagonal array of circular pixels, can result in a Sierpinski gasket pattern (Fig. 1b).

Another parameter is the rotation angle of the camera with respect to the monitor. Figure 1c shows a pattern recorded with a camera rotated by 45° and a magnification of $M \approx 1.29$. The experiment was greatly facilitated by our camera's ability to remove flicker-related effects by averaging over a number of frames. To our knowledge, this pattern, which arises directly from the pixellation of the display monitor, is the first published example of a stationary fractal created by unmodified video feedback.

Johannes Courtial, Jonathan Leach,
Miles J. Padgett

Department of Physics and Astronomy,
University of Glasgow, Glasgow G12 8QQ, UK
e-mail: j.courtial@physics.gla.ac.uk

1. Hofstadter, D. R. in *Gödel, Escher, Bach: an Eternal Golden Braid* 490–491 (Penguin, Harmondsworth, UK, 1980).
2. Andersen, M. C. The Ultimate Video Feedback Page, <http://www.videofeedback.dk/World/>
3. Cadman, J. Spinning Lights Home Page, <http://home.earthlink.net/~spinninglights/>
4. Crutchfield, J. P. *Physica D* **10**, 229–245 (1984).
5. Crutchfield, J. P. *IEEE Trans. Circuits Systems* **35**, 770–780 (1988).
6. Peitgen, H.-O., Jürgens, H. & Saupe, D. in *Chaos and Fractals: New Frontiers of Science* 19–22 (Springer, New York, 1992).
7. Peitgen, H.-O., Jürgens, H. & Saupe, D. in *Chaos and Fractals: New Frontiers of Science* 20 (Springer, New York, 1992).
8. King, P. H. Video Fractal Genesis, <http://www.sirius.com/~fringe/feedback/diagrams.html>
9. Courtial, J. & Padgett, M. J. *J. Mod. Opt.* **47**, 1469–1474 (2000).
10. Goodman-Strauss, C. Fractal Feedback! <http://comp.uark.edu/~cgstrauss/TV/index.html>
11. Peak, D. & Frame, M. in *Chaos Under Control* 301 (Freeman, New York, 1994).
12. Andersen, M. C. & Petersen, J. Simulation of Video Feedback, <http://www.videofeedback.dk/vf/ss/sspost.html> (1996).
13. Courtial, J. & Padgett, M. J. *Phys. Rev. Lett.* **85**, 5320–5323 (2000).

brief communications is intended to provide a forum for brief, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers (communications arising). Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com).

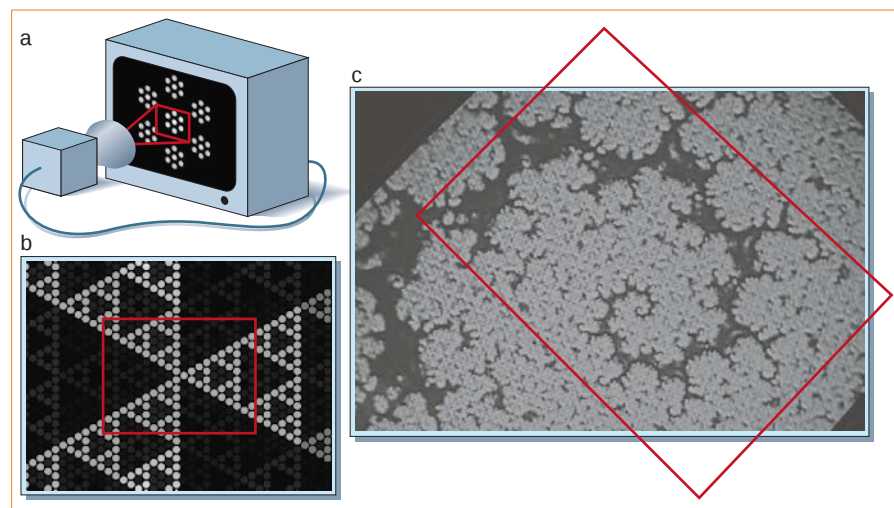


Figure 1 Stationary fractals resulting from pixellated video feedback. **a**, Basic set-up; the monitor displays an example of a self-similar pattern. **b**, Simulated video-feedback pattern consisting of pixel-limited Sierpinski gaskets. **c**, Pattern obtained with a rotated camera pointing at a colour monitor. In all three examples, the red rectangle marks the camera's respective field of view. Additional details can be obtained from the authors.

The DNA sequence and comparative analysis of human chromosome 20

P. Deloukas, L. H. Matthews, J. Ashurst, J. Burton, J. G. R. Gilbert, M. Jones, G. Stavrides, J. P. Almeida, A. K. Babbage, C. L. Bagguley, J. Bailey, K. F. Barlow, K. N. Bates, L. M. Beard, D. M. Beare, O. P. Beasley, C. P. Bird, S. E. Blakey, A. M. Bridgeman, A. J. Brown, D. Buck, W. Burrill, A. P. Butler, C. Carder, N. P. Carter, J. C. Chapman, M. Clamp, G. Clark, L. N. Clark, S. Y. Clark, C. M. Clee, S. Clegg, V. E. Cobley, R. E. Collier, R. Connor, N. R. Corby, A. Coulson, G. J. Coville, R. Deadman, P. Dhami, M. Dunn, A. G. Ellington, J. A. Frankland, A. Fraser, L. French, P. Garner, D. V. Grafham, C. Griffiths, M. N. D. Griffiths, R. Gwilliam, R. E. Hall, S. Hammond, J. L. Harley, P. D. Heath, S. Ho, J. L. Holden, P. J. Howden, E. Huckle, A. R. Hunt, S. E. Hunt, K. Jekosch, C. M. Johnson, D. Johnson, M. P. Kay, A. M. Kimberley, A. King, A. Knights, G. K. Laird, S. Lawlor, M. H. Lehtvaslaiho, M. Leversha, C. Lloyd, D. M. Lloyd, J. D. Lovell, V. L. Marsh, S. L. Martin, L. J. McConnachie, K. McLay, A. A. McMurray, S. Milne, D. Mistry, M. J. F. Moore, J. C. Mullikin, T. Nickerson, K. Oliver, A. Parker, R. Patel, T. A. V. Pearce, A. I. Peck, B. J. C. T. Phillimore, S. R. Prathalingam, R. W. Plumb, H. Ramsay, C. M. Rice, M. T. Ross, C. E. Scott, H. K. Sehra, R. Shownkeen, S. Sims, C. D. Skuce, M. L. Smith, C. Soderlund, C. A. Steward, J. E. Sulston, M. Swann, N. Sycamore, R. Taylor, L. Tee, D. W. Thomas, A. Thorpe, A. Tracey, A. C. Tromans, M. Vaudin, M. Wall, J. M. Wallis, S. L. Whitehead, P. Whittaker, D. L. Willey, L. Williams, S. A. Williams, L. Wilming, P. W. Wray, T. Hubbard, R. M. Durbin, D. R. Bentley, S. Beck & J. Rogers

The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

The finished sequence of human chromosome 20 comprises 59,187,298 base pairs (bp) and represents 99.4% of the euchromatic DNA. A single contig of 26 megabases (Mb) spans the entire short arm, and five contigs separated by gaps totalling 320 kb span the long arm of this metacentric chromosome. An additional 234,339 bp of sequence has been determined within the pericentromeric region of the long arm. We annotated 727 genes and 168 pseudogenes in the sequence. About 64% of these genes have a 5' and a 3' untranslated region and a complete open reading frame. Comparative analysis of the sequence of chromosome 20 to whole-genome shotgun-sequence data of two other vertebrates, the mouse *Mus musculus* and the puffer fish *Tetraodon nigroviridis*, provides an independent measure of the efficiency of gene annotation, and indicates that this analysis may account for more than 95% of all coding exons and almost all genes.

The finished reference sequence of the human genome is now in sight, underpinned by the recently published working draft^{1,2}. From the outset of the Human Genome Project, the plan has been to determine the complete sequence of each chromosome to an accuracy of greater than 99.99%, and to cover more than 95% of the gene-containing part of the genome (the euchromatin). This finished 'gold' standard was defined and upheld on completion of the first two human chromosomes, 22 and 21 respectively. Here we report completion of the sequence of the first metacentric human chromosome, chromosome 20, to these standards. Analysis of the finished sequence has benefited from comparison with substantial new data sets that were not available at the time of the previous finished chromosome analyses. These include new collections of human and mouse messenger RNA sequences, the protein indices of fully sequenced model organisms, and extensive sequencing of two vertebrates genomes, those of the mouse and the puffer fish *T. nigroviridis*. As a result, we were able to assess the quality and completeness of human gene annotation by independent analyses. The application of new analytical tools has also enabled assessment of predictive methods to define transcription start sites and other features of gene structures, although these require further development and calibration with the finished annotated sequence.

Clone map and finished sequence

We identified a set of 629 minimally overlapping clones (the tiling path) that spans the euchromatic regions of the short (p) and long (q) arm of human chromosome 20. The tiling path consists of 455 P1-derived artificial chromosomes (PACs), 169 bacterial artificial chromosomes (BACs), 3 yeast artificial chromosomes (YACs), 1 cosmid and 1 polymerase chain reaction (PCR) product (Fig. 1). The euchromatic portion of the chromosome is represented in six

contigs with one contig covering the entire p arm (Table 1). Boundaries between euchromatin and heterochromatin were identified by presence of satellite repeats in the sequence of clones located at the most distal and proximal ends, respectively, of the contigs flanking the centromere, and served as logical termination points for map construction. Clones located at the centromeric boundary of the p arm (Fig. 1) gave an additional signal at 20q11.1 upon fluorescent *in situ* hybridization (FISH) on metaphase chromosomes. We constructed an additional two-clone contig representing this duplication (Fig. 1) and postulate that it is located in the heterochromatic region of the q arm. In contrast to the p arm, four gaps remain in the clone map of the q arm. Three of them are clustered within a 1.2-Mb region at qtel (Fig. 1). We anticipate that the sequences in these gaps are unclonable to the host-vector systems used in this study, probably owing to the high guanine

Table 1 Sequence contigs on chromosome 20

Contig	Size (bp)	Size estimate (kb)
AL360078-AL358116	26,257,626	
Centromere		ND
AL121723-AL512784	5,063,606	
Gap		20
AL450465-AL450463	24,982,240	
Gap		~50
AL391316-AL499627	1,147,210	
Gap		~100
AL449263	35,826	
Gap		~150
AL450469-AL137028	1,700,790	
Total euchromatic sequence	59,187,298	
AL121762-AL441988	234,339	
Total sequence determined	59,421,637	
ND, not determined.		

and cytosine (G+C) content of the sequence in this region. All four euchromatic gaps were sized by FISH of clones immediately flanking each gap to extended DNA fibres. No gap was estimated to be larger than 150 kb and all the gaps together account for no more than 320 kb of DNA (Table 1). Finally, we defined the location of both telomeres. At the end of the p arm (ptel), clone RP11-530N10 (EMBL accession code AL360078; Fig. 1) ends about 10 kb away from the block of subtelomeric repeats, which extends for 40–50 kb on the basis of the telomeric half-YAC yRM2005 (ref. 3 and H. Riethman, personal communication). A larger allelic variant of the subtelomeric repeat block is also known, half-YAC yA35 (ref. 4). At the end of the q arm (qtel), clone RP11-476I15 (AL137028; Fig. 1) contains part of the subtelomeric repeat block. Each clone of the tiling path was subjected to random subcloning and sequencing. On the basis of internal and external⁵ quality checks, we estimate the accuracy of our finished sequence to exceed 99.99%. Each clone has been finished according to the agreed international finishing standard for the human genome (<http://genome.wustl.edu/gsc/Overview/finrules/hgfinrules.html>). In total, we finished 59,421,637 bases in seven sequence contigs. The size of each sequence contig is given in Table 1; the largest one spans the 26,257,626 bp of the entire p arm. The four gaps account for 0.32 Mb (Table 1). Thus, the sequence covers 99.46% of the euchromatic part of chromosome 20, which spans 59.5 Mb. Our estimate for the total size of the chromosome, based on size estimates of 3 Mb for the centromere and 0.2 Mb for subtelomeric repeats, is 62.7 Mb, which is smaller than a previous estimate of 72 Mb (ref. 6).

Gene index of chromosome 20

The finished genomic sequence was first analysed for G+C content and CpG islands. Interspersed and simple tandem repeats in the sequence were then masked and the masked sequence was compared against protein, DNA and expressed sequence tags (ESTs) using BLASTX and BLASTN⁷. In parallel, gene structures were predicted *ab initio* in the masked sequence on a clone-by-clone basis with the programs FGENESH⁸ and GENSCAN⁹.

A total of 895 gene structures was annotated in the finished sequence on the basis of human interpretation of the combined supportive evidence generated during sequence analysis (see Fig. 1). The structures were divided into five groups: (1) 335 'known' genes, that is, those that are identical to known human complementary DNA or protein sequences (all known genes were in the LocusLink database, <http://www.ncbi.nlm.nih.gov/LocusLink>); (2) 222 'novel genes', that is, those that have an open reading frame (ORF), are identical to human ESTs that splice into two or more exons, and/or have homology to known genes or proteins (all species); (3) 23 'novel transcripts', that is, genes as in 2 but for which a unique ORF cannot be determined; (4) 147 'putative genes', that is, sequences identical to human ESTs that splice into two or more exons but without an ORF; and (5) 168 'pseudogenes', that is, sequences homologous to known genes and proteins but with a disrupted ORF.

Excluding the pseudogenes, chromosome 20 has a gene density of 12.18 per Mb, which is intermediate to 6.71 (low) and 16.31 per Mb (high) reported for chromosome 21 and 22, respectively^{10,11}. We used the gene density of chromosomes 20, 21 and 22 from ref. 12 to adjust the number of genes on each of these chromosomes. The adjusted figures were then used to extrapolate a number of 31,500 genes for the whole genome, which is in agreement with recent estimates^{1,2}.

The analysis of chromosome 20 benefited from the availability of new large data sets to assist the gene annotation. These included human (for example, Genoscope) and mouse ESTs and 'full-length' cDNAs (for example, RIKEN mouse cDNA collection) as well as the protein indices of fully sequenced model organisms such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Arabidopsis thaliana*. Some 81% of the 557

genes in groups 1 and 2 (and 64% of all the annotated genes), have a full ORF as defined by a starting ATG codon and the presence of a 5' and a 3' untranslated region (UTR). Often a stretch of nucleotides immediately preceding the starting ATG seems to be part of the ORF. When the supporting evidence (for example, ESTs) terminated within such a stretch, we did not annotate a 5' UTR. Such genes were also included (8.9% of the 557 genes) in the above set.

The transcription start sites of most genes in the human genome are not yet known. We carried out several analyses to assist the annotation of the 5' ends of as many genes on chromosome 20 as possible. Analysis of the unmasked sequence predicted a total of 660 CpG islands, of which 389 are located near (5 kb upstream or 1 kb downstream) the first exon of an annotated gene structure. Many of the remaining predicted CpG islands have intragenic locations, which in our view does not allow a direct correlation between the observed number of CpG islands and the number of genes on chromosome 20. Among the genes with complete structures, 303 (67%) are associated with a CpG island at their 5' ends, which is in good agreement with the previously reported figure of 60% (ref. 13). We also scanned the sequence of chromosome 20 for putative transcription start (TS) sites using the probabilistic TS site detector program Eponine (T. Down, unpublished). Eponine is optimized for mammalian genomic DNA sequences and detects likely TS sites on the basis of the surrounding sequence (typically 500 bases upstream to 100 bases downstream). Multiple predictions are often clustered, suggesting alternative TS sites for a gene. Eponine has a detection sensitivity of 40%, on the basis of an analysis of human chromosome 22. We found 1,432 TS sites on chromosome 20, of which 492 (34%) are located within 2 kb of the first exon of an annotated gene. In the set of genes with complete structures, 402 TS sites are associated with 166 genes (37.5%) of which 159 (95.8%) have a CpG island at their 5' end. So, Eponine predicts multiple TS sites per gene (the mean value is 2.42 in the 166 genes) and has a bias in predicting TS sites in genes associated with a CpG island at their 5' end.

The 727 genes (that is, introns and exons) extend over a total of 25,213,914 bp (mean 34,682 bp per gene). Excluding expressed pseudogenes, 42.4% of the reported sequence of chromosome 20 is therefore transcribed. Exons account for only 2.43% of the sequence and the mean exon size is 283 bp. A summary per gene group is given in Table 2, which includes figures reported from the

Figure 1 The sequence map of human chromosome 20 and its features. The short (p) and the long (q) arm of the chromosome are depicted in the top and bottom panels, respectively. The features of each chromosome arm are shown from top to bottom as follows: (1) The finished sequence of each clone in the tiling path as a yellow line. Sequence positions are indicated in megabases along the x-axis of 'G+C content' (see 4, below). Eight of the clones were isolated and sequenced elsewhere, namely AC005808 (LBNL H136; BAC 185), AC005914 (LBNL H135; BAC 189), AC006076 (LBNL H133; PAC 12), AC004762 (LBNL H134; PAC 128), AC005220 (LBNL H80; BAC 99), AC004501 (LBNL H144; BAC 121) and AC004505 (LBNL H65; PAC 86C1) at the Joint Genome Institute¹⁶ and AC006198 (RP11-3A1) at the Whitehead Institute (Massachusetts Institute of Technology Center for Genome Research). The centromere has been arbitrarily drawn to span 3 Mb. The exact location of contig AL121762–AL441988 in the heterochromatic region of the q arm is not known. Gaps in the map appear as greenish bars. The width of the bar represents the size estimate obtained by fibre FISH. (2) The location of genetic markers. (3) The distribution of the main types of repeats in the sequence. (4) Plot of the G+C content of the sequence. (5) Plot of the SNP density along the sequence. (6) The location of predicted CpG islands. (7) The location of the annotated gene structures. Right and left coloured arrows indicate gene structures on the + and – strand, respectively. The most 3' end of each gene is drawn halfway along the arrowhead. Only the genes of the annotation group 1 (known; dark blue) and 2 (novel; blue) are named. When no gene symbol is available, the gene name used in the EMBL sequence submission file appears (for example, dJ583P15.4). CDS, protein-coding sequence.

Table 2 Structural characteristics of annotated gene structures

Chromosome, gene type	Mean size (kb)	Mean exon size (bp)	Mean number of exons
Chr20, known genes	51.3	294	10.3
Chr21, known genes	57.0		
Chr20, novel genes	25.1	278	5.7
Chr20, putative genes	9.1	217	2.5
Chr21, novel + putative genes	27.0		
Chr20, genes	34.7	283	7.1
Chr21, genes	39.0		
Chr20, pseudogenes	1.9	499	1.4
Chr20, all	27.6	292	6.0
Chr22, all	19.2	266	5.4

Structural characteristics of groups of annotated gene structures on chromosome 20 are shown, and compared with similar groups on chromosomes 21 and 22.

analyses of chromosomes 21 and 22 (refs 10, 11). Gene size varies substantially, from 1,234,386 bp (gene C20orf133 (AL117333–AL049633)), which is similar to a low-density lipoprotein-related protein, LRP16) to 339 bp (gene C20orf127 (AL121753)). Exon sizes are fairly constant, with the exception of 3' terminal exons (for example, 8,181 bp in *PTPRT*), in contrast to intron sizes, which vary from 33 bp (C20orf97 (AL034548)) to 523,790 bp (*CDH4*).

For 209 (29%) of the annotated genes, we found alternative splice forms. Alternative splicing can, for example, give rise to two distinct peptides by exclusion of the exons encoding a functional domain from one but not the other transcript. The transcript for the soluble form of attractin (*ATRN*; AL353193–AL132773) lacks the five exons that encode the transmembrane and cytoplasmic domains and are present in the transcript that encodes the membrane form of the protein. Splice variants may encode structurally unrelated peptides. A complex example of alternative splicing and genetic imprinting is found in the *GNAS1* locus (AL132655–AL109840). *NESP55* and *XLAS* are transcribed from distinct mono-allelic promoters located upstream of the bi-allelic promoter that drives the transcription of the gene for the α -subunit of the stimulatory guanine-nucleotide-binding protein G_s . $G_{s\alpha}$ is encoded by exons 1–13. A large G protein, $XL\alpha_s$, is generated by in-frame splicing of an upstream exon (bp 120,789–121,953; AL132655) to exon 2 (bp 39,075–39,119; AL121917) of $G_{s\alpha}$. An additional exon located further upstream (bp 106,368–107,508; AL132655) splices again to exon 2 of $G_{s\alpha}$ but not in frame, giving rise to a structurally unrelated peptide, *NESP55*. An antisense transcript (dJ806M20.3.6; AL132655) postulated to regulate this imprinted region has also been reported¹⁴. In total, we annotated six isoforms of *GNAS1*. One gene (*PLCB4*; AL121898–AL031652) was found to have the most isoforms

(eight); in most cases of genes with alternative splicing (130 genes) we observed two isoforms. Typically, we annotated the longest possible terminal exon and did not create entries for alternative splice forms on the basis of alternative polyadenylation sites. If we exclude the putative genes that have mainly incomplete structures, then 35% of the genes (average of 1.65 transcripts per gene) show alternative splicing. This is in agreement with previous estimates¹⁵. Analysis of chromosomes 19 and 22 (ref. 1), both gene-rich chromosomes, showed a higher extent of alternative splicing.

Protein index of chromosome 20

We analysed the proteome of chromosome 20 using InterProScan (<http://www.ebi.ac.uk/interpro/scan.html>) to look at the distribution of known protein domains. The InterPro database combines information on protein families, domains and functional sites from the databases Pfam, PRINTS, PROSITE, SMART and SWISS-PROT (see <http://www.ebi.ac.uk/interpro> for links). Of all proteins encoded on chromosome 20, 73.5% have an InterPro match and 30% are multidomain with an average of 2.1 distinct InterPro domains. As shown in Table 3, many of the most frequent domains in the chromosome 20 proteome rank in similar order as in the human proteome¹. There are, however, five domains for which chromosome 20 seems enriched. Four of them—the cysteine proteases inhibitor (IPR000010), the immunoglobulin subtype (IPR003599), the whey acidic protein (WAP)-type 'four-disulphide core' domain (IPR00222), and the pancreatic trypsin inhibitor (Kunitz/Bovine) domain (IPR002223)—are found in proteins encoded by three gene clusters along the chromosome.

Functionally related gene clusters indicate probable ancestral gene-duplication events. The first cluster, at 1.5 Mb (Fig. 1), extends from AL109658 to AL034562 and includes genes with immunoglobulin and immunoglobulin-like domains that are involved in signal transduction and cell adhesion (*SIRPB1*, *SIRPB2* and *PTPNS1*). The annotated genes *PTPN1L* and *PTPNS1L2* and pseudogenes dJ576H24.1 and dJ673D20.1 are new members of this gene family. Interestingly, the apparently functional gene *PTPNS1L2* is located within a larger fragment of 33,048 bp (AL049634 and AL592544), which is an insertion type of polymorphism. The insertion allele is represented in the RP4 PAC library but not the RP11 BAC library. Using a panel of 174 Caucasians, we estimated that the frequency of the insertion allele is 37.3%. The second cluster, at 23.5 Mb (AL096677–AL121831; Fig. 1), comprises members of the cystatin gene family, which encode protease inhibitors with antibacterial and antiviral activities. An additional member, *CST7*, is located about 1 Mb distal of the main

Table 3 Most common InterPro domains in the chromosome 20 proteome and their abundance in other species

Rank (genome rank)	InterPro code	Abundance						Name
		Chr20	Hs	Dm	Ce	At	Sc	
1 (2)	IPR000822	26	576	341	205	169	53	Zinc finger, C2H2 type
2 (3)	IPR000719	12	481	230	419	1,033	116	Eukaryotic protein kinase
	(IPR002290)	12	316	158	219	856	113	Serine/threonine protein kinase family active site)
	(IPR001245)	11	193	82	121	477	4	Tyrosine kinase catalytic domain)
3 (ND)	IPR002965	11	190	175	62	178	0	Proline rich extension
4 (ND)	IPR000010	9	20	4	3	7	0	Cysteine proteases inhibitor
	(IPR003243)	8	10	0	1	7	0	Cystatin C and M)
5 (4)	IPR000276	9	373	84	368	0	0	Rhodopsin-like GPCR superfamily
6 (13)	IPR000561	9	234	86	137	42	1	EGF-like domain
7 (24)	IPR000008	8	111	42	53	99	11	C2 domain
8 (ND)	IPR000504	8	203	135	111	248	54	RNA-binding region RNP-1 (RNA recognition motif)
9 (1)	IPR003006	8	584	134	66	0	0	Immunoglobulin and major histocompatibility complex domain
	(IPR003599)	7	177	27	11	0	0	Immunoglobulin subtype)
10 (ND)	IPR000345	7	2	4	3	3	3	Cytochrome c family haem-binding site
11 (ND)	IPR001124	7	8	0	10	2	0	Lipid-binding serum glycoprotein
12 (ND)	IPR002221	7	6	4	7	0	0	WAP-type four-disulphide core domain
13 (ND)	IPR002223	6	13	22	38	0	0	Pancreatic trypsin inhibitor (Kunitz/Bovine) family

At the time of analysis, the InterPro database contained 18,149 *Homo sapiens* (Hs, incomplete), 13,843 *Drosophila melanogaster* (Dm), 18,581 *Caenorhabditis elegans* (Ce), 25,677 *Arabidopsis thaliana* (At) and 6,176 *Saccharomyces cerevisiae* (Sc) protein entries. Rows in parentheses correspond to InterPro domains that are 'children' (= members) of a broader InterPro domain. ND, not determined. The P-loop motif domain IPR001687 was excluded, owing to low specificity.

cluster (AL035661). Only two of the known cystatins are not on chromosome 20 (IPR003243; Table 3). We annotated three new members (*CST8L*, *CSTL1* and *CST9L*) and two pseudogenes. The third cluster, at 43.5 Mb (AL049767–AL050348; Fig. 1), includes 11 genes that encode proteins with a WAP-type four-disulphide core domain (IPR002221) and/or a pancreatic trypsin inhibitor (Kunitz/Bovine) domain (IPR002223). A fourth cluster that includes genes for the semenogelins SEMG1 and SEMG2 (semen proteins involved in reproduction) is located within the third gene cluster between members *PI3* and *SLPI*, which have only a WAP-type domain.

Chromosome landscape

The sequence of chromosome 20 has an average G+C content of 44.1%, which is slightly higher than the genome average of 41%. The distribution of the G+C content fluctuates along the chromosome, and regions with higher G+C have a higher gene density (Fig. 1). For example, the sequence from 49.5 to 54 Mb has an average G+C content of 41.1% and a gene density of only 4.9 genes per Mb, in contrast to the region between 60 and 62.5 Mb, which has 56.6% G+C content and a gene density of 28 genes per Mb. Given a sequence length and gene ratio of 1.25 and 1.65, respectively, between the q and the p arm, the q arm seems rich in genes. Gene density can drop as low as 1.54, for instance in a 1.9-Mb region between AL136990 and AL139163. Interestingly, the largest genes, such as *PTPRT*, *PLCBI* and *dj631M13.5* are located adjacent to or within gene-poor regions.

The repeat content of chromosome 20 is 42%. The distribution of the main classes of repeats (detailed in Supplementary Information) is shown in Fig. 1. Regions of high gene density seem enriched in short interspersed elements (SINEs).

Segmental duplications are another interesting feature of the genome. We compared the masked sequence of chromosome 20 with the rest of the genome and with itself to identify inter- and intrachromosomal duplications, respectively. The segments of chromosome 20 involved in interchromosomal duplications (Fig. 2) often contain pseudogenes; for example, at 6 Mb, AL359954 contains a pseudogene similar to *TRDBP* that maps to 1p36 (AL109811). The region at 53.9 Mb (Fig. 2) that is duplicated in chromosomes 21 and 22 was recently described as part of a breast cancer amplicon¹⁶. A region of about 500 kb between 25.8 and 26.3 Mb is implicated in both types of segmental duplications. A core region of 100 kb that harbours a copy of exon 7 of the *CFTR* gene is duplicated on chromosome 20. The second copy is located in AL121762–AL441988 at the pericentromeric region of the q arm. Copies of the extended region seem to be present on chromosomes 9, 12, 15, 17 and 19 (Fig. 2). Secondary signals in the pericentromeric regions of these and other chromosomes were also observed on FISH analysis of clones AL078587 and AL121762. It will be interesting to investigate whether the gene structures annotated in AL121762 and AL441988 are expressed genes, particularly C20orf80, which is similar to the *FRG1* gene. *FRG1* is located 100 kb centromeric of the repeat units on chromosome 4q35, which are deleted in facioscapulohumeral muscular dystrophy. The region from 48.2 to 48.8 Mb is bordered by two copies of a 60-kb intrachromosomal duplication.

The integrated Marshfield male, female and sex-averaged genetic maps of chromosome 20 (ref. 17) were aligned to the physical map (Fig. 3). The steepest increase in recombination frequency is observed between markers D20S178 and D20S176, which are both located in the region of duplication described above. A

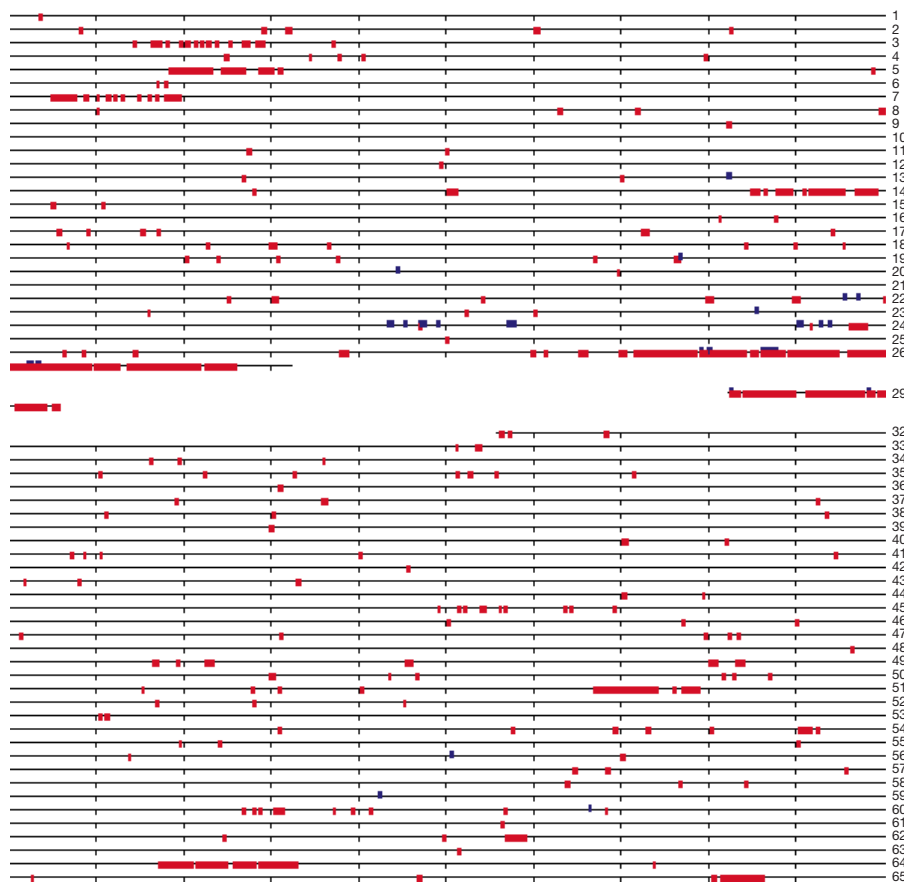


Figure 2 Duplication landscape of chromosome 20. Intrachromosomal and inter-chromosomal duplications are shown in blue and red, respectively. Each horizontal line represents 1 Mb of the sequence from the telomeric end of the short arm (top left) to the

telomeric end of the long arm (bottom right). The gap indicates the centromeric region. Pairwise alignments generated by Exonerate and longer than 1 kb are shown.

region of very low recombination extends for about 20 Mb between markers D20S432 and D20S859. The rate of recombination in specific loci differs between the two sexes. Compared with that of the female, the rate of male recombination is higher along the p arm up to marker D20S432 and lower across the rest of the chromosome (Fig. 3).

Sequence variation

The definition of the common ancestral haplotypes that are present in the population relies on the availability of an extensive collection of single nucleotide polymorphisms (SNPs). We first placed 26,678 SNPs (deposited in the dbSNP database, <http://www.ncbi.nlm.nih.gov/SNP>) on the sequence of chromosome 20. Of those, 13,016 were derived from sequence analysis of clone overlaps by the program ssahaSNP¹⁸. To recover additional SNPs in clone overlaps, we realigned all available clone-based shotgun sequences from chromosome 20 (including unfinished sequence in clone overlaps that was previously archived and therefore excluded from the earlier analysis) onto the finished sequence with ssahaSNP, and detected 11,050 SNPs (submitted to dbSNP). Merging the two data sets resulted in 32,763 unique SNPs on chromosome 20 (Fig. 1), of which 6,085 are new. In the unique set, there are 14,211 SNPs (43.4%) located within annotated genes and 3,061 of them are in exons.

Comparative analysis

Functional features such as exons and regulatory elements have been conserved through evolution and there is compelling evidence that comparative genomic sequence analysis is a powerful tool in the quest to complete the structural annotation of the human genome. Two data sets were available in the public domain at the time of analysis: about 13 million sequence reads of a mouse whole-genome shotgun giving an estimated genome coverage of 2.3-fold (<http://trace.ensembl.org>), released by the Mouse Sequencing Consortium on 8 May 2001; and 816,262 single sequence reads from BAC and plasmid ends of the *T. nigroviridis* genome, totalling 663,839,518 bases and corresponding to 1.72 genome equivalents, generated at Genoscope. Thus we undertook the comparative analysis of the finished and annotated sequence of an entire human chromosome against two vertebrate genomes.

Mouse sequences were aligned to the sequence of chromosome 20 using Exonerate version 0.3d (Guy StC. Slater, unpublished). We obtained matches with 63,644 mouse sequences representing 12,041 regions of sequence conservation (RSC) along chromosome 20. *Tetraodon* sequences were aligned at Genoscope by Exofish ('exon finding by sequence homology'), which generates ecores (evolutionary conserved regions)¹⁹. Matches were obtained with 2,992 ecores (available at <http://www.genoscope.cns.fr/exofish>). We first examined the annotated gene structures; 77.4% of the 727 genes and 89% of the 168 pseudogenes have at least one exon matched by a mouse RSC or *Tetraodon* ecore. This figure is much higher for the 557 genes in groups 1 and 2 (94%) than it is for the 'putative' gene structures in group 4 (33%). Furthermore, the two sets differ in the ratio of genes with only a mouse RSC to genes with both an RSC and ecore match: 1:3.8 and 1:0.2, respectively. These observations suggest that the putative gene structures may represent largely UTRs, which have sequences known to be less well conserved between species, and possibly genes that appeared later in evolution.

We then looked at matches outside annotated exons as a way to assess the completeness of the current annotation. Such matches may correspond to exonic sequences that have not been annotated in the present study owing to lack of supporting evidence (for example, EST, cDNA and protein homologies). Note that we did not use RSCs and ecores during the annotation process. We found 5,447 RSC and 207 ecore matches, and 60 of these non-exonic regions are conserved in all three species. Of all annotated exons (including pseudogenes), 2,050 (36.3%) contain a region conserved in all three

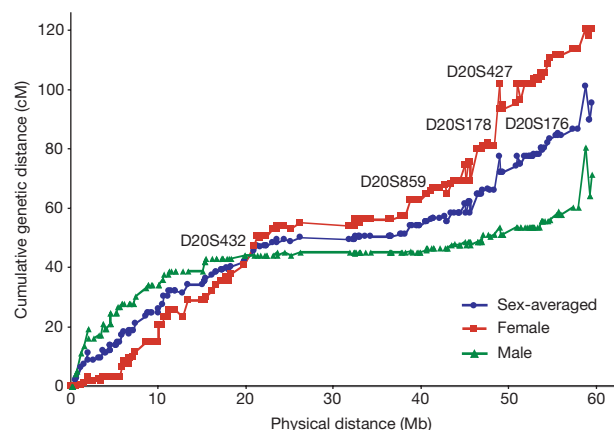


Figure 3 Alignment of the genetic map of chromosome 20 to the physical map. The two maps are aligned from the telomeric end of the short arm to the telomeric end of the long arm. The position of each genetic marker on the female, the male and the sex-averaged genetic map is indicated.

species (in contrast to 0.2% of the annotated introns). Thus, we postulate that about 97.2% ($2,050 / (2,050 + 60)$) of all coding exons of chromosome 20 have been annotated in this study. A caveat is whether the set of annotated genes used in this analysis is representative of genes that appeared recently in evolution, as ecores are biased to more conserved genes; however, we consider that such an effect cannot be substantial.

The 639 and 4,808 RSC matches in annotated introns and intergenic regions, respectively, suggest that although the mouse data set provides better coverage (70% of all exons) than the ecores, exonic sequences cannot be readily identified by simple comparison at the DNA level.

GENSCAN can be used on small segments of genomic sequence to effectively evaluate the likelihood of that segment containing an exon. We performed a GENSCAN analysis on 'extended RSCs', which included 100 bp of human sequence either side of the RSC match, to divide them into those that were more likely to be coding regions of sequence conservation (cRSC) and those more likely to be noncoding. This predicted 3,299 cRSCs and 8,836 noncoding RSCs and found that 65.7% of cRSCs match annotated exons (3.8% are within introns). As a result, the 874 cRSCs found between annotated genes is a set enriched in regions that may represent non-annotated exons (there are 4,748 RSC matches in intergenic regions).

Conclusion and medical implications

We sequenced the euchromatic portion of human chromosome 20 leaving four small gaps that account for no more than 320 kb. In the 59,421,637 bp of sequence, we annotated 727 gene structures of which 64% are complete and 168 pseudogenes. A comparison of this product with the draft assembly of chromosome 20 reported earlier this year² clearly shows the importance of generating a contiguous finished reference sequence for each human chromosome. Both the G+C and gene density plots of chromosome 20 peak between 60 and 62.5 Mb (Fig. 1) at the qtel region, which is in sharp contrast to the corresponding plots shown in Fig. 11 in ref. 2, which peak at least 12 Mb proximal of the telomere. Furthermore, the order in which genes are shown in the magnified part of Fig. 13 in ref. 2 is incorrect. For example, the gene *OSBPL2* (oxysterol binding protein 2) and *BB379O24.1* (*GATA5* related) are located at 60.2–60.6 Mb and cannot map between *PTPRT* (protein tyrosine phosphatase, receptor type) at 41 Mb and *ZNF217* (Kruppel-like transcription factor) at 51.7 Mb. The use of the clone map information was instrumental in resolving similar problems during the assembly of the chromosome 20 draft sequence.

The output of the comparative analysis of chromosome 20 from the mouse whole-genome shotgun and the ecores generated from the *Tetraodon* genomic sequence suggests that the current sets of human and mouse ESTs and 'full-length' cDNAs together with the proteomes of model organisms are adequate to allow the identification of the vast majority of human genes in the sequence. As expected, we found that comparative analysis can be used to reliably identify exonic sequences. The mouse shotgun data alone cannot be used reliably to postulate the number of non-annotated exons, owing to the overall higher degree of sequence conservation. The use of the two data sets together, however, provides an excellent tool for assisting the identification of new, and the completion of existing, gene structures. In the present study, the ability to identify regulatory elements in the sequence of chromosome 20 by comparison to the mouse sequence data can be substantiated only by anecdotal evidence. A three-way comparison with the addition of the genome sequence of a species more closely related to humans may hold the key in this endeavour²⁰.

Chromosome 20 is best known for harbouring the genes that cause Creutzfeldt–Jakob disease (*PRNP*) and severe combined immunodeficiency (*ADA*). However, the causes of the sporadic cases of Creutzfeldt–Jakob disease (80% of all cases) remain unknown, and no mutation in the *ADA* gene has been identified to explain the phenotype of *ADA* excess in haemolytic anaemia. Furthermore, there are still single-gene disorders mapped to chromosome 20 (<http://www.ncbi.nlm.nih.gov/Omim>) for which the underlying genetic defect is not known. The resources generated by the Human Genome Project have already been used to accelerate the cloning of disease genes on chromosome 20; the Alagille (*JAG1*)²¹, McKusick–Kaufman (*MKKS*)²², ICF (*DNMT3B*)²³ and Hallervorden–Spatz (*PANK2*)²⁴ syndromes are recent examples. The reported finished and annotated sequence and its variation will be a valuable tool in tackling not only the remaining single-gene diseases but also the multifactorial diseases that have been linked to chromosome 20, such as type 2 diabetes, obesity, cataract, eczema and Grave's disease. Evidence for a susceptibility locus for hereditary prostate cancer on 20q13 has also been reported^{25,26}. In addition to the sequence itself, the isolated clones used in the sequencing process constitute a unique resource in studying chromosome loss and/or amplification in various types of cancer. We have recently reported the refinement of a commonly deleted region (CDR) of 20q12–13.1 found in patients with myeloproliferative disorders and myelodysplastic syndromes²⁷. Others have reported the characterization of a breast cancer amplicon at 20q13.2 (ref. 16), whereas several studies have reported loss of heterozygosity across regions of 20q using comparative genomic hybridization^{28,29}. □

Methods

Clone map and sequence assembly

Clone map construction is described in ref. 30. Mapped sequence tagged sites (STSs) for screening genomic PAC and BAC libraries were selected from the integrated radiation hybrid map constructed for chromosome 20 (<http://www.sanger.ac.uk/cgi-bin/rhmap?chr=20>), which harbours 1,493 STS-based markers. In regions with no clone coverage, screening was extended to the CEPH (Centre d'Etude du Polymorphisme Humain), ICRF (Imperial Cancer Research Fund) and ICI (Imperial Chemical Industries) YAC libraries and the LANL (Los Alamos National Laboratory) chromosome-20-specific cosmid library (links for the libraries can be found at http://www.hgmp.mrc.ac.uk/Biology/descriptions/genomic_libraries.html). For the shotgun phase, pUC plasmids with inserts of 1.4–2 kb were sequenced from both ends by the dideoxy chain termination method³¹ with big dye terminator chemistry³². Most of the reactions were analysed on ABI3700 capillary sequencing machines. The resulting data were processed by a suite of in-house programs (<http://www.sanger.ac.uk/Software/sequencing>) before assembly with the PHRED^{33,34} and PHRAP (<http://www.phrap.org>) algorithms. For the finishing phase, we used the GAP4 program³⁵ to help assess, edit and select reactions, eliminate ambiguities and close sequence gaps. Sequence gaps were closed by a combination of primer walking, PCR, short/long insert sublibraries³⁶, sublibrary screening with oligonucleotides and, in rare cases, transposon sublibraries.

Sequence analysis tools

Interspersed and simple tandem repeats were identified with Repeatmasker (<http://repeatmasker.genome.washington.edu>) and etandem (<http://www.emboss.org>),

respectively. BLAST 1.4 (default parameters and matrix; <http://blast.wustl.edu>) was used to identify initial matches, which were then re-aligned by EST_GENOME³⁷. BLASTN was used with a 65% similarity cutoff in the comparison against the RIKEN mouse cDNA set³⁸ instead of 95%, which is used when searching human ESTs, to find significant matches. In the unmasked sequence, CpG islands were predicted by searching for sequence segments that are at least 400 bp, have a G+C content greater than 50%, and an expected/observed CpG count of greater than 0.6. The completed analysis was assembled into contigs and visualized in AceDB (<http://www.acedb.org>), whereas an Ensembl (<http://www.ensembl.org>) database of the sequence assembly and the annotated genes was constructed and used for calculation of statistics and producing Fig. 1. In SNP analysis, only those regions of chromosome 20 where a SNP was detected by at least four reads was considered valid, since the depth of shotgun sequencing for these clones was greater than 4×. The phred-quality value of at least four of the reads at the SNP location had to be at least 30 (error probability of phred base calling 0.001 or less). Exonerate was run with an initial word length of 14 bp, gap penalties of 8 for opening a gap and 4 for extending one, and a score of 5 and –4 for DNA matches and mismatches, respectively.

Received 13 August; accepted 25 October 2001.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, C. J. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
4. Chute, L. L., Ashley, T. & Dobson, M. J. The telomere-associated DNA from human chromosome 20p contains a pseudotelomere structure and shares sequences with the subtelomeric regions of 4q and 18p. *Genomics* **41**, 51–60 (1997).
5. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
6. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
7. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
8. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).
9. Burge, C. & Karlin, S. Prediction of gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
10. Hattori, M. *et al.* The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature* **405**, 311–319 (2000).
11. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
12. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
13. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA* **90**, 11995–11999 (1993).
14. Hayward, B. E. & Bonthron, D. T. An imprinted antisense transcript at the human *GNAS1* locus. *Hum. Mol. Genet.* **9**, 835–841 (2000).
15. Brett, D. *et al.* EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.* **474**, 83–86 (2000).
16. Collins, C. *et al.* Comprehensive genome sequence analysis of a breast cancer amplicon. *Genome Res.* **11**, 1034–1042 (2001).
17. Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
18. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: A fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
19. Roest Crolius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
20. Dubchak, I. *et al.* Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* **10**, 1304–1306 (2000).
21. Li, L. *et al.* Allagile syndrome is caused by mutations in human *Jagged1*, which encodes a ligand for Notch1. *Nature Genet.* **16**, 243–251 (1997).
22. Stone, D. L. *et al.* Mutation of a gene encoding a putative chaperonin causes McKusick–Kaufman syndrome. *Nature Genet.* **25**, 79–82 (2000).
23. Hansen, R. S. *et al.* The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl Acad. Sci. USA* **96**, 14412–14417 (1999).
24. Zhou, B. *et al.* A novel pantothenate kinase gene (*PANKS*) is defective in Hallervorden–Spatz syndrome. *Nature Genet.* **28**, 345–349 (2001).
25. Berry, R. *et al.* Evidence for a prostate cancer-susceptibility locus on chromosome 20. *Am. J. Hum. Genet.* **67**, 82–91 (2000).
26. Zheng, S. L. *et al.* Evidence for a prostate cancer linkage to chromosome 20 in 159 hereditary prostate cancer families. *Hum. Genet.* **108**, 430–435 (2001).
27. Bench, A. *et al.* Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig and identification of candidate genes. *Oncogene* **19**, 3902–3913 (2000).
28. Bench, A. *et al.* A detailed physical and transcriptional map of the region of chromosome 20 that is deleted in myeloproliferative disorders and refinement of the common deleted region. *Genomics* **49**, 351–362 (1998).
29. Rigaud, G. *et al.* High resolution allelotyping of nonfunctional pancreatic endocrine tumors: Identification of two molecular subgroups with clinical implications. *Cancer Res.* **61**, 285–292 (2001).
30. Bentley, D. R. *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* **409**, 942–943 (2001).
31. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
32. Rosenblum, B. B. *et al.* New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**, 4500–4504 (1997).
33. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).

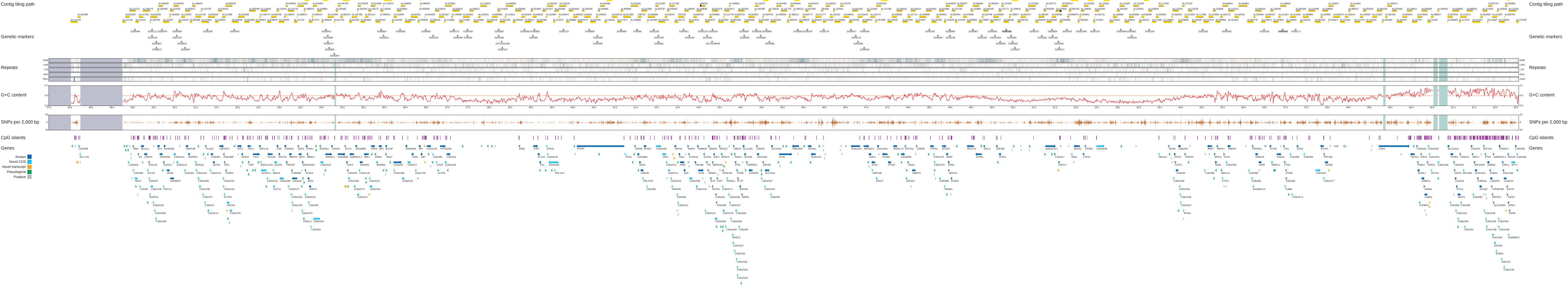
34. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
35. Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**, 4992–4999 (1995).
36. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**, 562–566 (1998).
37. Mott, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* **13**, 477–478 (1997).
38. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685–690 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

All the reported DNA sequences have been deposited to the EMBL, GenBank or DDBJ databases. We thank H. Roest Crollius, O. Jaillon and the *Tetraodon* group at Genoscope for the Exofish data; E. Zdobnov for the InterPro analysis; the Ensembl team for assistance in the computational analysis; and the group of M. S. Jackson at the University of Newcastle upon Tyne for markers and clones that assisted the mapping of the pericentromeric duplication. Finally, we thank H. Wain, E. Bruford, M. Lush, R. Lovering, M. Wright, P. Shetty and S. Povey for help with gene nomenclature and the Wellcome Trust for financial support.

Correspondence and requests for material should be addressed to P.D. (e-mail: panos@sanger.ac.uk).



Structural basis of water-specific transport through the AQP1 water channel

Haixin Sui^{*†}, Bong-Gyoon Han^{*†}, John K. Lee^{*‡}, Peter Walian^{*} & Bing K. Jap^{*‡}

^{*} Life Sciences Division, Lawrence Berkeley National Laboratory; and [‡] Graduate Group in Comparative Biochemistry, University of California, Berkeley, California 94720, USA

[†] These authors contributed equally to this work

Water channels facilitate the rapid transport of water across cell membranes in response to osmotic gradients. These channels are believed to be involved in many physiological processes that include renal water conservation, neuro-homeostasis, digestion, regulation of body temperature and reproduction. Members of the water channel superfamily have been found in a range of cell types from bacteria to human. In mammals, there are currently 10 families of water channels, referred to as aquaporins (AQP): AQP0–AQP9. Here we report the structure of the aquaporin 1 (AQP1) water channel to 2.2 Å resolution. The channel consists of three topological elements, an extracellular and a cytoplasmic vestibule connected by an extended narrow pore or selectivity filter. Within the selectivity filter, four bound waters are localized along three hydrophilic nodes, which punctuate an otherwise extremely hydrophobic pore segment. This unusual combination of a long hydrophobic pore and a minimal number of solute binding sites facilitates rapid water transport. Residues of the constriction region, in particular histidine 182, which is conserved among all known water-specific channels, are critical in establishing water specificity. Our analysis of the AQP1 pore also indicates that the transport of protons through this channel is highly energetically unfavourable.

AQP0–AQP9 can be divided into two major groups^{1,2}: AQP0–AQP2, AQP4–AQP6 and AQP8, permeable to water but not to small organic and inorganic molecules³, and AQP3, AQP7 and AQP9, permeable to glycerol or urea as well as water^{1,3}. Both groups of channels are active in many physiological functions^{4,5}.

AQP1 (relative molecular mass, $M_r = 28,000$) was initially found in red blood cells and renal proximal tubules^{5,6}. AQP1 water channels allow water, but not ions including protons, to move freely and bidirectionally across the cell membrane^{3,7}. Sequence analysis shows high homology among members of the AQP1 family and that the two halves of the sequence exhibit a high degree of similarity². Each sequence half contains an NPA (asparagine, proline, alanine) motif, which is conserved throughout the AQP superfamily, including the glycerol facilitators. Moderate-resolution projection and low-resolution 3D (three-dimensional) maps of AQP1 derived from electron crystallographic studies provided the first structure-based evidence for a general architecture consisting of six helices surrounding two putative helical structures within the membrane bilayer^{8–11}. Models of AQP1 derived from electron crystallographic structural studies at about 4 Å resolution have recently been reported; they independently confirm the presence of two membrane-inserted non-membrane-spanning helices^{12–14}.

We report here the structure of AQP1 from bovine red blood cells at 2.2 Å resolution as determined by X-ray crystallography (Table 1). At this resolution the positions of the side chains that establish the properties of the transmembrane channel pathway, as well as water molecules captured in transit, are clear. Critical differences exist between this AQP1 structure and those determined by electron crystallography, particularly in the pore profiles^{12–14}. For example, the constriction region in models derived from these previous studies is approximately 8 Å away from this region in the model of AQP1 presented here. The constriction region in our model is also much smaller than that inferred from the high-resolution structure of the *Escherichia coli* glycerol facilitator, GlpF¹⁵. The high-resolution structure of AQP1 now reveals the structural basis for water specificity; together with the structure of GlpF it also

provides the molecular details necessary to understand the mechanisms regulating water and other solute selectivity within the aquaporin superfamily.

Architectural overview

The functional unit of AQP1 is a tetramer with each monomer providing an independent water pore (Fig. 1a, c). Each monomer contains six transmembrane helices packed to form part of a trapezoid-like structure when viewed normal to the membrane

Table 1 Data collection and refinement statistics

Data set	Resolution (Å)	Complete (%)	R_{sym} (%)	I/σ_I
Au-AQP1	2.2 (2.30–2.20)	99.7 (100.0)	6.2 (58.2%*)	41.8 (6.3)
TI-AQP1				
λ1 (0.97871)	2.8 (2.85–2.80)	99.2 (99.8)	5.5 (41.8%)	50.0 (9.7)
λ2 (0.97954)	2.8 (2.85–2.80)	99.2 (100.0)	5.0 (29.2%)	53.5 (13.8)
λ3 (1.00800)	2.8 (2.85–2.80)	99.4 (100.0)	5.1 (30.2%)	54.5 (13.6)
λ4 (0.95593)	2.8 (2.85–2.80)	99.2 (99.8)	6.6 (47.9%)	45.8 (8.3)
Refinement statistics				
Resolution (Å)	15–2.2			
Reflections in working/test set	16,934/1,235			
R (%)	26.6			
R_{free} (%)	30.8			
Nonglycine residues in most favourable region of Ramachandran plot (%)	97.2			
Outliers and generously allowed regions (%)	2.8			
r.m.s. deviation from ideality				
Bond lengths (Å)	0.0066			
Bond angles (degrees)	1.25			

Au-AQP1 indicates the data set collected from crystals grown in the presence of gold cyanide and TI-AQP1 denotes the data obtained from thallium-derivatized AQP1 crystals. The labels λ1–λ4 indicate the different data sets collected for multi-wavelength anomalous diffraction (MAD) phasing. The wavelengths used are shown in parentheses.

* This large R_{sym} value is the consequence of significant anisotropy. When the data were filtered at a 1σ level, the percentage of reflections removed from the quadrants of maximum anisotropy was proportionally the highest and the following values were obtained: $R_{\text{sym}} = 34.9\%$, completeness was 78.0% and $I/\sigma_I = 6.7$.

$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the measurement of intensity. $R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ for all values and R_{free} is the R value for the reflections that were excluded in refinement, where F_{obs} is the observed native amplitude and F_{calc} is the one calculated from atomic models. The values for the highest resolution zone are shown in parentheses.

plane. Two membrane-inserted but non-membrane-spanning helices, which define a major portion of the pore, are partially enclosed by this structure and form an integral part of the outer wall as well. This general folding topology is similar to that reported for the structure of *E. coli* GlpF¹⁵ and to those reported from electron crystallographic studies of AQP1 (refs 12–14).

Residues comprising the six transmembrane and the two membrane-inserted non-membrane-spanning helices (M1–M8) are depicted in Fig. 1b. The amino terminus of the molecule is located at the cytoplasmic side of the cell membrane, as determined from earlier biochemical studies^{16,17}, and leads into the first of two transmembrane helices. These two helices (M1, M2) are followed by a membrane-inserted loop that contains an NPA motif and leads

into the non-membrane-spanning helix (M3). The loop exiting this very short helix enters the cytoplasm and turns back into the membrane to connect to transmembrane helix M4, which ends the N-terminal half of the structure on the extracellular side of the cell membrane. The second half of the structure is essentially an inverted repeat of the topology of the first half and results in the location of the carboxy terminus at the cytoplasmic face. Three nonylglucoside detergent molecules have been located on a region of the monomer surface in contact with the extracellular leaflet of the lipid bilayer. Each monomer is ~40 Å across and ~60 Å long.

Structure of the pore

The pore of AQP1 is dumbbell-like in shape when viewed in profile

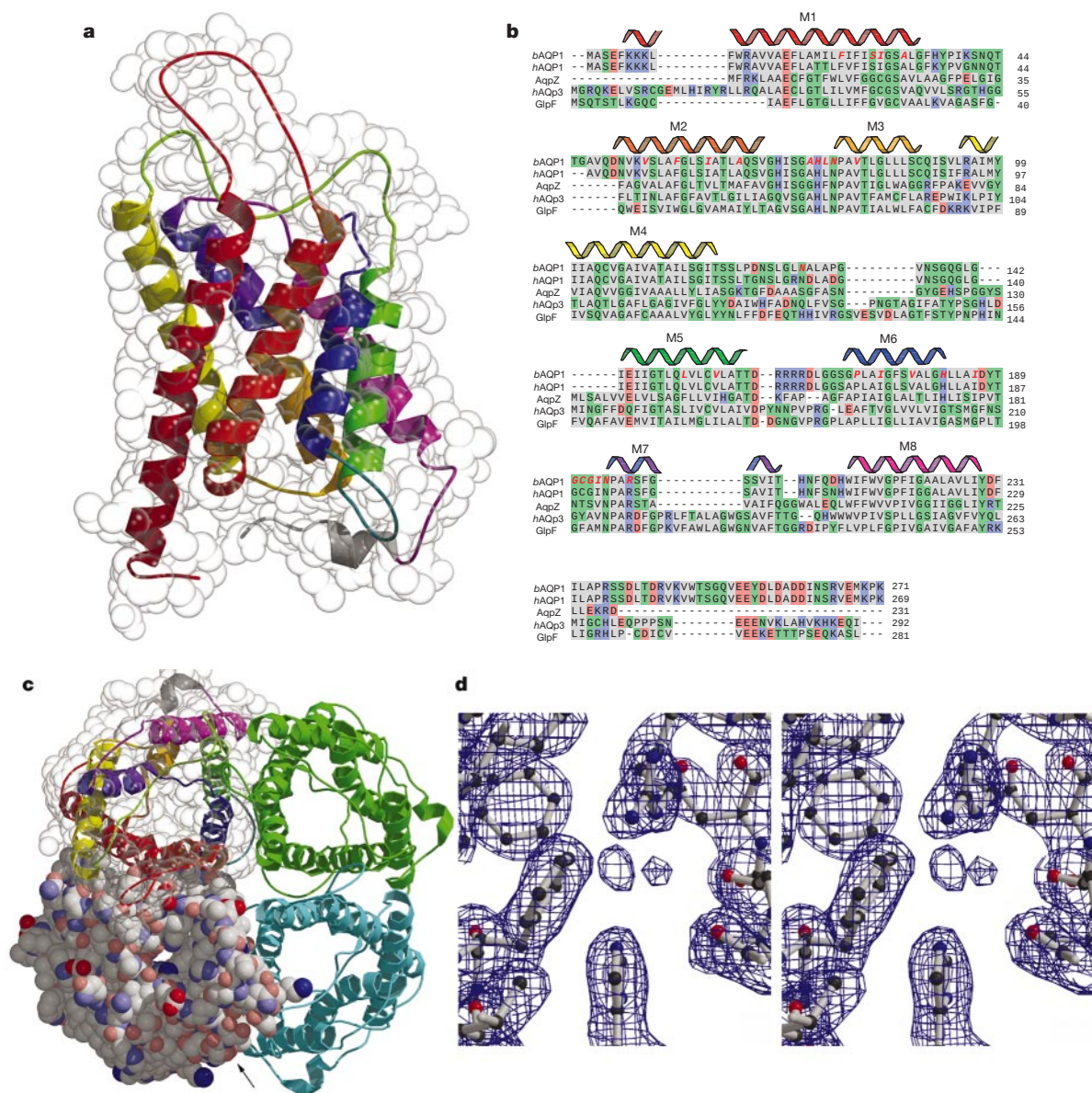


Figure 1 Models of AQP1, sequence alignment of selected superfamily members and a view of the density map. **a**, Combined ribbon diagram and space-filling model of AQP1 monomer viewed parallel to the membrane. **b**, Amino-acid sequence alignment of bovine AQP1, human AQP1, *E. coli* AqpZ, human AQP3 and *E. coli* GlpF produced using CLUSTAL W5. AQP1 membrane-embedded helices are colour-coded as shown in **a** and labelled M1–M8. M3 and M7 are membrane-inserted non-membrane-spanning helices. Residues lining the extended narrow pore (selectivity filter) of bovine AQP1 are indicated in red italicized letters. **c**, AQP1 tetramer viewed normal to the membrane; two of the

monomers also contain a space-filling model to show the pore entrance. The black arrow points along the direction of the cut-line used to produce the monomer view in Fig. 2b. **d**, Stereo view of the 2Fo – Fc electron density map (1.5σ contour) and the corresponding region of the AQP1 model. The view is from the extracellular side, down the pore, and centred about the constriction region. The spherical densities near the pore centre are those of the two water molecules located in the extracellular half of the selectivity filter. The figure was produced using MOLSCRIPT³⁶ and Raster3D³⁷.

(Fig. 2a, b), consisting of three general elements, an extracellular vestibule, an extended narrow pore or selectivity filter containing the constriction region, and a cytoplasmic vestibule. Helices M4 and M8 do not contribute residues to the pore. About half of the channel wall along the selectivity filter can be considered hydrophobic and the other half hydrophilic. The hydrophilic face provides the chemical groups that are essential for displacing certain waters of hydration and therefore establishes a pathway for coordinating water transport.

The steep crossing angles of the monomer helices aid in the formation of the extracellular and cytoplasmic vestibules. Shaping of the vestibule mouths is completed by loop regions at each monomer face and, in addition, by the N- and C-terminal residues at the cytoplasmic face (Fig. 1a, c). The extracellular vestibule is roughly conical in shape with a mouth diameter of about 15 Å (pore size estimates are based on the van der Waals radii used in the AMBER program suite¹⁸). The population of polar moieties along the surface of the extracellular vestibule consists predominantly of polar residues, only a small number of which are charged, and polar groups from the solvent-exposed backbone of extended loop regions.

Over a distance of about 20 Å, the extracellular vestibule tapers down to its narrowest point, ~2.8 Å in diameter, forming the constriction region and the beginning of the ~20-Å-long selectivity filter (Fig. 3). A series of solvent-accessible carbonyl oxygens forms a path leading from the extracellular vestibule through the constriction regions of both AQP1 and GlpF; in the case of bovine AQP1 these groups are provided by residues G190, C191, G192 and I193 of

the connecting loop leading into non-transmembrane helix M7 (Fig. 4). This architectural scheme, which positions a helix linker loop so as to form a key element of the selectivity filter, is reminiscent of the structural motif employed in the potassium channel from *Streptomyces lividans* (KcsA) for the removal of solute hydration waters¹⁹.

At the constriction region, residues H182 and R197, along with the solvent-accessible carbonyl oxygen of residue C191, form the hydrophilic face of the pore (Fig. 5). The imidazole ring of H182 is fully extended into the pore while the bulk of R197 is pointed upwards almost parallel to the pore axis in a manner similar to that observed for the equivalent arginine of the GlpF channel¹⁵. Opposite the hydrophilic face at the constriction region is the hydrophobic face defined here by residue F58.

Three of the four residues defining the constriction region of the AQP1 pore (R197, H182 and F58) are conserved across the water-specific aquaporins²⁰. Access to the carbonyl group of the fourth residue, C191, seems to be essential at that location. The conservation of arginine, histidine and phenylalanine side chains at their respective locations within the constriction region in the known water channels is a strong indicator of channel water specificity and may be useful in assigning function to the sequences of candidate aquaporins for which physiological assays have not been performed.

In GlpF the H182 of AQP1 is replaced by glycine. This substitution provides the additional room needed to accommodate a second substitution, phenylalanine for C191 (Fig. 5). These substitutions have two critical effects on the characteristics of the GlpF constriction region: they increase its size and its hydrophobicity.

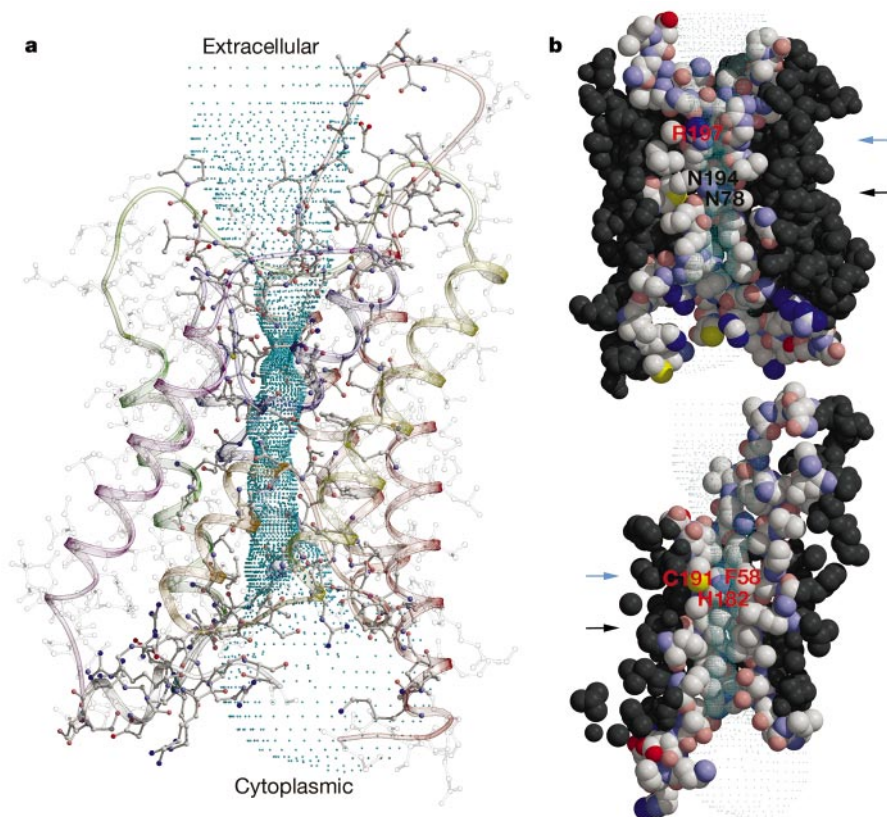


Figure 2 Side-views of AQP1. **a**, Backbone in ribbon format, residues depicted in ball-and-stick representation; residues lining the pore are shown in opaque colours. The pore profile is highlighted by an array of blue dots generated by the program HOLE³⁸. The constriction region is visible as the pinched-in area in the extracellular half of the profile. **b**, Side-view cutaway of a space-filling model made along the axis indicated by the black arrow in Fig. 1c. Residues not in direct contact with the pore are dark grey. The pore

lining in the top figure consists primarily of polar groups distributed along the length of the selectivity filter; the complementary half is predominantly hydrophobic in nature. The constriction region and pseudo two-fold axis (centred about the NPA motifs) are indicated by blue and black arrows respectively. The pore profile (blue dots) is superimposed on each channel half. The figure was produced using MOLSCRIPT³⁶ and Raster3D³⁷.

Past the constriction region, the pore opens up again, averaging about 4 Å in diameter over the next ~15 Å. About 8 Å past the constriction point, residues from the two highly conserved NPA motifs are brought into close proximity owing to the end-to-end packing of the two short membrane-inserted non-membrane-spanning helices M3 and M7. This places the terminal amine groups of the two NPA motif asparagine residues, N194 and N78, in the pore. The pseudo two-fold axis of the molecule runs parallel to the membrane plane and is located approximately halfway between these two short helices.

As in the case of GlpF, short helices M3 and M7 are situated end-to-end and lengthwise along one side of the channel so that the positive ends of their helical dipoles point inward toward the pore. The aquaporins therefore differ from KcsA in their use of such non-membrane-spanning helices; in KcsA, four such helices are positioned radially around the pore axis and the direction of the dipoles is reversed, establishing a net negative zone in the middle of the pore¹⁹. The two-helix aquaporin motif is found in both water- and glycerol-selective channels, which suggests that this structure is not used in discriminating between water and glycerol. The examples of KcsA and the aquaporins suggest that pore-located non-membrane-spanning helices may serve as functionally critical structural motifs in other classes of transport-associated membrane proteins.

A second 'string' of carbonyl oxygens lines the pore, this time extending away from the NPA motifs towards the cytoplasmic vestibule and are provided by residues (L77, H76, A75 and G74) of the connecting loop from M2 into the other membrane-inserted non-membrane-spanning helix, M3 (Fig. 4). The complete set of pore-accessible carbonyl oxygens and asparagine amine groups are

arranged in a long-pitched helical pattern forming much of the hydrophilic half of the selectivity filter (Figs 2b and 4). A similar distribution of pore-lining groups was reported for the structure of GlpF (ref. 15).

Near the cytoplasmic end of the selectivity filter is another pore-accessible histidine, H76. Unlike H182, this histidine is highly conserved across the aquaporin superfamily and glycerol facilitators. Instead of extending maximally into the pore, as does the constriction defining H182, H76 is fixed against the side of the pore wall. The GlpF counterpart to H76 is oriented within the pore in a similar fashion.

In the last 8–10 Å of the channel the pore flares out to form a cytoplasmic vestibule with a mouth approximately 15 Å wide. As its walls are somewhat more uniform in height, this vestibule is more conical in shape than the extracellular vestibule, and its mouth is essentially perpendicular to the pore axis. Here the concentration of polar residues increases significantly from that occurring in the selectivity filter; these residues establish a zone within the inner half of this vestibule that is more hydrophilic than that of GlpF (Fig. 3).

Location of waters in the channel

Waters have been identified at four locations within the AQP1 selectivity filter (Fig. 4). The density attributable to a single water molecule is located adjacent to the constriction region about 7 Å from the pseudo two-fold axis towards the extracellular surface. At this location water is coordinated by hydrogen bonds established with the ε2 nitrogen of H182 and the backbone carbonyl oxygen of G192. The next two waters are centred about the pseudo two-fold

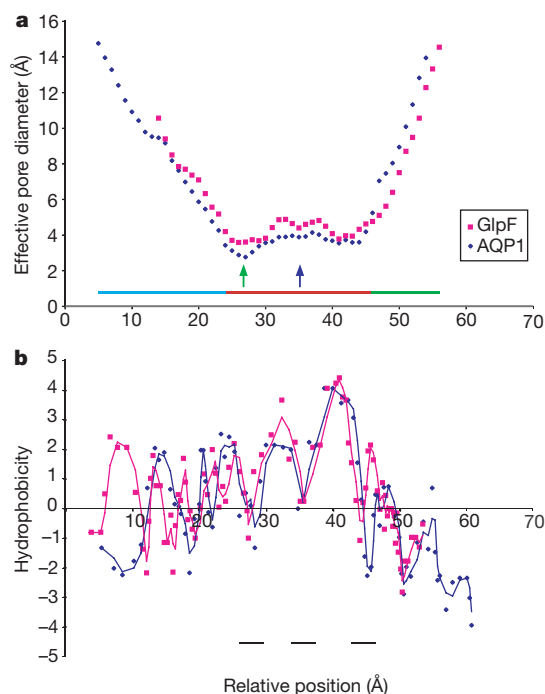


Figure 3 The effective pore diameter (**a**) and hydrophobicity (**b**) of the AQP1 and GlpF channels. Green and dark blue arrows indicate the locations of the constriction region and the pseudo two-fold axis respectively. Below these arrows a light blue bar indicates the location of the extracellular vestibule, a red bar the selectivity filter and a green bar the cytoplasmic vestibule. The three black bars in the hydrophobicity profile identify hydrophilic nodes within the selectivity filter. Pore diameters were determined with AMBER-based van der Waals radii¹⁸ and analysed using the program HOLE³⁸. Hydrophobicity was characterized using the Kyte and Doolittle amino-acid hydropathy scale³⁹ and a four-point averaging window across the pore-lining residues.

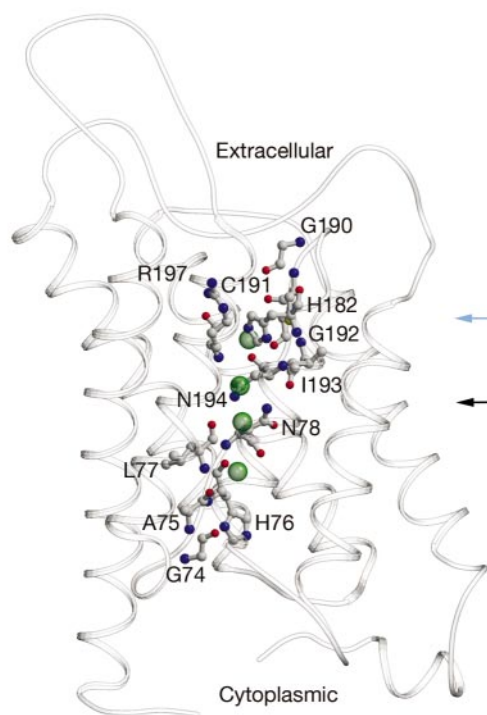


Figure 4 Selectivity filter water molecules and residues forming the hydrophilic face of the channel pore. Cut-away side view of the channel with secondary structure shown in ribbon format. Side chains critical to establishing the long-pitched helical, hydrophilic path across the length of the selectivity filter are shown. The four water molecules located within the selectivity filter are depicted as green spheres. Of these four water molecules only the middle two are close enough to form a water–water hydrogen bond. The constriction region and pseudo two-fold axis (centred about the NPA motifs) are indicated by light blue and black arrows respectively. The figure was produced using MOLSCRIPT³⁶ and Raster3D³⁷.

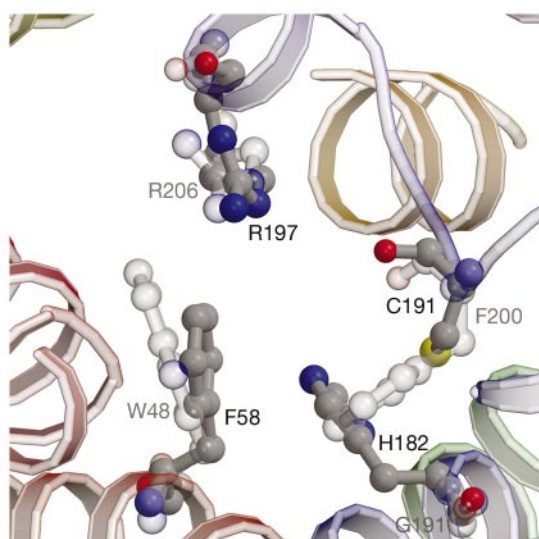


Figure 5 Residues defining the constriction region. Residues involved in the formation of the AQP1 constriction region (H182, R197, F58 and C191) are depicted in solid colours while the transparent side chains are those of the equivalent residues found in the structure of GlpF (G191, R206, W48 and F200). These side chain differences result in a larger and more hydrophobic constriction region in GlpF. Alignment of the residues was performed using a least-square fit of their backbone atoms. The figure was produced using MOLSCRIPT³⁶ and Raster3D³⁷.

axis, one nearest and hydrogen-bonded to the $\delta 2$ nitrogen of N194 and the other nearest and hydrogen-bonded to nitrogen $\delta 2$ of N78. The fourth water visible in the channel is located near the cytoplasmic end of the selectivity filter. The backbone carbonyl oxygens of residues H76 and A75 form the coordinating hydrogen bonds with this water. These four waters do not form a contiguous hydrogen-bonded chain as only the middle two are close enough to each other to form a water–water hydrogen bond. The hydrophobicity profile of the residues lining the pore indicates that there are three hydrophilic nodes distributed along the length of the selectivity filter (Fig. 3). As would be expected, the four waters identified within the selectivity filter are all located at these nodes.

Mechanisms of water selectivity

Through a combination of channel sterics and solute binding sites, AQP1 facilitates the rapid and highly selective throughput of water. A steric limit of ~ 2.8 Å is established at the constriction region, and the chemical properties of the residues forming this structure provide additional criteria for solute selection. From the steric limit alone, it is now clear why the transport of glycerol by AQP1 is highly unfavourable. For a molecule of water to diffuse across the narrow AQP1 constriction region, its effective diameter must be reduced by shedding waters of hydration. For this to be an energetically favourable process, interactions with primary hydration shell waters must be replaced by interactions with residues lining the channel wall over as small a diffusion distance as possible. In AQP1, sufficient hydrogen-bond-forming groups are available so that water molecules can readily move through the constriction region. These bond-forming groups are provided by constriction region residues H182 (conserved across the water-specific aquaporins) and R197 (conserved throughout most of the aquaporin superfamily), and the backbone carbonyl oxygens of residues G190, C191 and G192.

In addition to the bound water found at the constriction region, there are three additional waters in the selectivity filter. Averaging only about 4 Å in diameter, the selectivity filter is also rather hydrophobic but is punctuated by water-binding regions at

several hydrophilic nodes (Fig. 3). The availability of water-binding sites at these nodes reduces the energy barrier to water transport across this predominantly hydrophobic pathway, while the relatively low number of such sites keeps the degree of solute–pore interaction to a minimum. In balancing these opposing factors the aquaporins are able to transport water selectively while optimizing permeability.

Whereas the aquaporins are optimized for the rapid transport of water, some members of the aquaporin family also facilitate the transport of other solutes. The bacterial aquaglyceroporin homologue, GlpF, is optimized for the rapid transport of glycerol. Although water has favourable steric accessibility through both the AQP1 and GlpF channels, there is a critical difference in glycerol steric accessibility between these two pores; the constriction region of GlpF is almost 1 Å wider than in AQP1. The large difference in glycerol permeability between AQP1 and GlpF is effectively bridged in aquaglyceroporin AQP3, which transports both water and glycerol at moderate rates²¹. The hydrophathy profiles for the selectivity-filter regions of AQP1, AQP3 and GlpF are very similar. Most of the amino-acid differences occurring in this region are moderate, such as swapping one type of hydrophobic residue for another; yet, changes to only one or two residues of the constriction regions in these channels are sufficient to radically alter solute selectivity. It appears that in AQP1 the residue most critical in supporting rapid water throughput, while hindering the passage of glycerol, is H182 (Fig. 5). Throughout the aquaporin family the choice of amino acid at this location appears to be essential in defining whether an aquaporin will be specific for water or additionally selective for other solutes such as glycerol. In GlpF, the replacement of this histidine by glycine serves both to significantly increase the size of the constriction region and sterically to support a second residue change, C191 to phenylalanine. These changes in turn alter the polar nature of this area and result in improved channel interactions with the hydrophobic backbone of glycerol. H182 of AQP1 is probably also replaced by glycine in AQP3. As with GlpF, such a substitution should provide the additional room needed to accommodate the side chain of a second probable substitution, in this case a tyrosine for C191. In contrast to the substitution for phenylalanine occurring at this position in GlpF, the switch to tyrosine potentially provides an additional location for solute hydrogen bonding. As a consequence of these two substitutions (H182G, C191Y) it is expected that the constriction region of AQP3 is similar in size to that of GlpF, although somewhat more polar, thereby allowing for the moderately rapid transport of glycerol as well as water.

The amphipathic nature of the selectivity filter has a key role in the rapid transport of water in AQP1, and this property is preserved across the superfamily of aquaporins. It has been shown to be particularly beneficial for glycerol transport. The hydrophobic face of the GlpF selectivity filter provides an ideal match for the carbon-backbone side of glycerol while the opposing hydrophilic face is equipped with hydrogen-bonding groups to replace waters of hydration in the vicinity of glycerol hydroxyl groups¹⁵. Both glycerol and water molecules were visualized within the GlpF selectivity filter, indicating that at high concentrations of glycerol there should be almost one-to-one transport of glycerol and water. Although GlpF permeability to water is significant, it is substantially less than that measured for its bacterial aquaporin counterpart and AQP1 homologue, AqpZ (ref. 22). The constriction region of GlpF is wider than that of AQP1, so it would appear that the more hydrophobic nature of GlpF is the primary factor responsible for its relatively reduced ability to transport water.

Preselection of solutes by the vestibules may also be important in maximizing rates of transport. The relatively stronger hydrophobic nature of the vestibules in GlpF should improve the transport rates of solutes that contain a significant hydrophobic component, such as glycerol, although potentially at the expense of rapid water

throughput. Conversely, the more hydrophilic vestibules of AQP1 favour the preselection of water.

Barriers to ion transport

In the aquaporin and KcsA K^+ channel selectivity filters, carbonyl groups extending from inner helix linker loops face the pore staggered along the channel axis. In AQP1, and as reported in the structure of GlpF¹⁵, these ten pore-accessible carbonyl oxygen atoms are distributed along a ~ 25 -Å stretch along one side of the pore. For KcsA this zone is significantly shorter, confined to half of the transmembrane pore, and yet contains 16 carbonyl oxygen atoms arranged in a stacked ring configuration, with four carbonyls per ring¹⁹; hydrated K^+ ions are closely and symmetrically surrounded by carbonyl oxygen atoms, so that making coordination site transfers between hydrating water and the carbonyl oxygens is energetically favourable. Hydrated ions encountering the type of selectivity filter utilized by the aquaporins would find that the available carbonyl groups and water-coordinating residues (such as R197 and H182 in AQP1) could only effectively substitute for less than half of the hydrating waters at any point along this segment of the channel. Such partially hydrated ions would still be too large to pass through the constriction region or most of the selectivity filter. Positively charged ions would find additional resistance to their transport from repulsive forces generated through interactions with the dipoles of helices M3 and M7 (oriented with their positive ends pointed into the pore), the highly conserved R197 (located near the constriction region), the amines of N78 and N194 (located about the pseudo two-fold axis) and by the histidines H182 and H76 located at opposite ends of the selectivity filter. Negatively charged ions would experience repulsive forces from the many pore-accessible carbonyl oxygen atoms lining the selectivity filter and inner regions of the vestibules.

Water networks or chains have been seen in the high-resolution structures of proteins such as the photosynthetic reaction centre²³ and cytochrome b_6f (ref. 24) and it has been suggested that protons could be transported along a continuous linear network of suitably oriented hydrogen-bonded waters ('proton-wire') by a transfer process termed the Grotthuss mechanism^{25,26}. In this mechanism a proton can be 'instantaneously' shuttled along this special network of water molecules. However, no suitable chain of hydrogen-bonded water molecules spanning the selectivity filter has been located in the density map of the AQP1 crystal structure. Factors preventing the establishment of such a hydrogen-bonded water chain along the string of carbonyl oxygen atoms lining the AQP1 channel include the following highly conserved elements: arginine located at the extracellular side of the selectivity filter, the two asparagines and helical dipole end charges located in the pore about the pseudo two-fold axis^{13,14}, and the histidines located at the ends of the selectivity filter. The two positive helical dipole end charges are by themselves sufficient to disrupt any suitably oriented alignment of water dipoles along the channel pore. □

Methods

Crystallization of AQP1

Native crystals were obtained as described previously²⁷. Soaking of crystals in a variety of heavy-atom compounds greatly affected diffraction quality even for very low concentrations of heavy atom compounds. Co-crystallization yielded thallium-derivatized crystals suitable for providing initial phasing information.

Data collection and processing

MAD (four wavelength) data sets were collected from Tl-AQP1 crystals at the Advanced Light Source (ALS) of Lawrence Berkeley National Laboratory. Diffraction data were processed with DENZO/SCALEPACK²⁸. Crystals belong to space group I422 (unit cell dimensions are: $a = b = 93.3$, $c = 180.5$ Å, $\alpha = \beta = \gamma = 90^\circ$) and contain one molecule per asymmetric unit. The best crystals diffracted to 2.1 Å. A Patterson map based on dispersive and anomalous differences showed strong heavy-atom peaks located near the crystal's four-fold symmetry axis. These positions were refined with the use of the SHARP program²⁹. The refined heavy-atom positions were used to obtain an initial density map.

An improved density map was produced with the application of solvent flattening through the program DM³⁰ and was used to build an initial model at 2.8 Å resolution.

Model building

Model building was performed using program O³¹ while model refinement was conducted using the Crystallographic and NMR system (CNS)³². Extension of phases and model refinement to 2.2 Å was accomplished with a data set obtained from crystals grown in the presence of gold cyanide but for which no multi-wavelength anomalous differences were observed. Others have also obtained improvements in crystal order using gold cyanide as a crystallization additive³³. The diffraction data from these crystals extended to 2.1 Å but were significantly anisotropic, which resulted in a reduction in the percentage of statistically reliable data in the direction of the h - k plane. CNS-based refinement of the model (residues 1–249 of a possible 271) using data to 2.2 Å resolution resulted in an improved model with an R_{free} value of 30.8% and a crystallographic R of 26.6%. The $2F_o - F_c$ electron density map obtained from these data is, however, of good quality (Fig. 1d). The relatively high values of R and R_{free} are a consequence of the diffraction set anisotropy; such an effect was expected and larger R factors have been reported for anisotropic data sets of similar resolution which also yielded good quality maps as reported in ref. 34.

Received 11 July; accepted 30 October 2001.

- Ishibashi, K. *et al.* Cloning and functional expression of a new aquaporin (AQP9) abundantly expressed in the peripheral leukocytes permeable to water and urea, but not to glycerol. *Biochem. Biophys. Res. Commun.* **244**, 268–274 (1998).
- Borgnia, M. *et al.* Cellular and molecular biology of the aquaporin water channels. *Annu. Rev. Biochem.* **68**, 425–458 (1999).
- Meinild, A. K. *et al.* Bidirectional water fluxes and specificity for small hydrophilic molecules in aquaporins 0–5. *J. Biol. Chem.* **273**, 32446–32451 (1998).
- Verkman, A. S. Water channels in cell membranes. *Annu. Rev. Physiol.* **54**, 97–108 (1992).
- Agre, P. *et al.* Aquaporin CHIP: the archetypal molecular water channel. *Am. J. Physiol.* **265**, F463–F476 (1993).
- Denker, B. M. *et al.* Identification, purification, and partial characterization of a novel M, 28,000 integral membrane protein from erythrocytes and renal tubules. *J. Biol. Chem.* **263**, 15634–15642 (1988).
- Zeidel, M. L. *et al.* Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes. *Biochemistry* **33**, 1606–1615 (1994).
- Jap, B. K. & Li, H.-L. Structure of osmo-regulated H_2O channel, AQP-CHIP, in projection at 3.5 Å resolution. *J. Mol. Biol.* **251**, 413–420 (1995).
- Li, H.-L., Lee, S. & Jap, B. K. Molecular design of aquaporin-1 water channel as revealed by electron crystallography. *Nature Struct. Biol.* **4**, 263–265 (1997).
- Cheng, A. *et al.* Three-dimensional organization of a human water channel. *Nature* **387**, 627–630 (1997).
- Wälz, T. *et al.* The three-dimensional structure of aquaporin-1. *Nature* **387**, 624–627 (1997).
- Ren, G. *et al.* Three-dimensional fold of the human AQP1 water channel determined at 4 Å resolution by electron crystallography of two-dimensional crystals embedded in ice. *J. Mol. Biol.* **301**, 369–387 (2000).
- Murata, K. *et al.* Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
- Ren, G. *et al.* Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl Acad. Sci. USA* **98**, 1398–1403 (2001).
- Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
- Smith, B. L. & Agre, P. Erythrocyte M, 28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins. *J. Biol. Chem.* **266**, 6407–6415 (1991).
- Nielsen, S. *et al.* Distribution of the aquaporin CHIP in secretory and absorptive epithelia and capillary endothelia. *Proc. Natl Acad. Sci. USA* **90**, 7275–7279 (1993).
- Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Park, J. H. & Saier, M. H. Jr Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J. Membr. Biol.* **153**, 171–180 (1996).
- Zeuthen, T. & Klaerke, D. A. Transport of water and glycerol in aquaporin 3 is gated by H^+ . *J. Biol. Chem.* **274**, 21631–21636 (1999).
- Borgnia, M. J. & Agre, P. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **98**, 2888–2893 (2001).
- Baciu, L. & Michel, H. Interruption of the water chain in the reaction center from *Rhodobacter sphaeroides* reduces the rates of the proton uptake and of the second electron transfer to Q_B . *Biochemistry* **34**, 7967–7972 (1995).
- Ponamarev, M. V. & Cramer, W. A. Perturbation of the internal water chain in cytochrome f of oxygenic photosynthesis: loss of the concerted reduction of cytochromes f and b_L . *Biochemistry* **37**, 17199–17208 (1998).
- Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **224**, 456–462 (1995).
- Pomès, R. & Roux, B. Free energy profiles for H^+ conduction along hydrogen-bonded chains of water molecules. *Biophys. J.* **75**, 33–40 (1998).
- Sui, H. *et al.* Crystallization and preliminary X-ray crystallographic analysis of water channel AQP1. *Acta Crystallogr. D* **56**, 1198–1200 (2000).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Cowan, K. D. An automated procedure for phase improvement by density modification. Joint CCP4 ESF-EACBM Newsletter. *Protein Crystallogr.* **31**, 34–38 (1994).

31. Jones, T. A. *et al.* Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
32. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
33. Chang, G. *et al.* Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* **282**, 2220–2226 (1998).
34. Groll, M. *et al.* The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc. Natl Acad. Sci. USA* **96**, 10976–10983 (1999).
35. Thompson, J. D. *et al.* CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
36. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
37. Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
38. Smart, O. S. *et al.* The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
39. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

Acknowledgements

Preliminary data collection and screening for heavy-atom derivatives were conducted at beamlines X25 at the National Synchrotron Light Source and 1–5 at Stanford Synchrotron Radiation Laboratory; data sets used in determining the model were collected at Beamline 5.0.2, Advanced Light Source, Lawrence Berkeley National Laboratory. We would like to thank all staff members of these beamlines for their assistance and B.-C. Wang for discussions on heavy-atom position refinement. This work is supported by funding from the National Institutes of Health and by the Office of Health and Environmental Research, US Department of Energy. The coordinates have been deposited in the Protein Data Bank under accession number 1J4N.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.K.J. (e-mail: BKJap@lbl.gov) or P.W. (e-mail: PJWalian@lbl.gov).

Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth

George Cooper*, Novelle Kimmich*, Warren Belisle*, Josh Sarinana*, Katrina Brabham* & Laurence Garrel†

* NASA Ames Research Center, Moffett Field, California 94035, USA

† International Research School of Planetary Sciences (IRSPS), Università d'Annunzio Viale Pindaro, 42, 65127 Pescara, Italy

The much-studied Murchison meteorite is generally used as the standard reference for organic compounds in extraterrestrial material. Amino acids and other organic compounds¹ important in contemporary biochemistry are thought to have been delivered to the early Earth by asteroids and comets, where they may have played a role in the origin of life^{2–4}. Polyhydroxylated compounds (polyols) such as sugars, sugar alcohols and sugar acids are vital to all known lifeforms—they are components of nucleic acids (RNA, DNA), cell membranes and also act as energy sources. But there

has hitherto been no conclusive evidence for the existence of polyols in meteorites, leaving a gap in our understanding of the origins of biologically important organic compounds on Earth. Here we report that a variety of polyols are present in, and indigenous to, the Murchison and Murray meteorites in amounts comparable to amino acids. Analyses of water extracts indicate that extraterrestrial processes including photolysis and formaldehyde chemistry could account for the observed compounds. We conclude from this that polyols were present on the early Earth and therefore at least available for incorporation into the first forms of life.

Our study of Murchison and Murray polyols uses a relatively definitive means of detection, gas chromatography–mass spectrometry (GC–MS). We identified all compounds as their trimethylsilyl (TMS) and/or tertiary butyl-dimethylsilyl (*t*-BDMS) derivatives (see Methods). Earlier studies^{5,6} on the possible presence of sugars in (pre-Murchison) meteorites had reported the six-carbon (6C) sugars glucose and/or mannose, and much smaller amounts of the five-carbon (5C) sugars arabinose and xylose. These earlier studies had been done on a variety of carbonaceous and non-carbonaceous meteorites, using thin-layer chromatography and ultraviolet–visible, infrared, and nuclear magnetic resonance spectroscopy. Most of the examined meteorites had fallen several

	Sugars	Sugar Alcohols	Sugar Acids	Dicarboxylic Sugar Acids	Deoxy Sugar Acids			
3C	CH₂OH C=O CH₂OH Dihydroxyacetone	CH₂OH H—C—OH CH₂OH Glycerol 160 nmol/g (100%)	CO₂H H—C—OH CH₂OH Glyceric acid 80 nmol/g	—				
4C	—	CH₂OH H—C—OH H—C—OH CH₂OH Erythritol & Threitol (1%)	CO₂H H—C—OH H—C—OH CH₂OH Erythronic & Threonic acid (4nmol/g)	CO₂H H—C—OH HO—C—H CO₂H Tartaric & Mesotartaric acid	CO₂H H₃C—C—OH CH₂OH 2-Methyl glyceric acid	CO₂H H—C—OH H—C—H CH₂OH 2, 4 Dihydroxy butyric acid	CO₂H H—C—OH H—C—OH CH₃ 2, 3 Dihydroxy butyric acid (& diastereomer)	CO₂H H—C—H H—C—OH H—C—OH CH₂OH 3, 4 Dihydroxy butyric acid
5C	—	CH₂OH H—C—OH H—C—OH H—C—OH CH₂OH Ribitol & Isomers	CO₂H H—C—OH H—C—OH H—C—OH CH₂OH Ribonic acid & Isomers	CO₂H H—C—OH HO—C—H H—C—OH CO₂H 2, 3, 4-Trihydroxy Pentanedioic acid	CO₂H H—C—H H—C—OH H—C—OH CH₂OH 2-Deoxypentonic acids			
6C	*	CH₂OH H—C—OH HO—C—H H—C—OH H—C—OH CH₂OH Glucitol & Isomers	CO₂H H—C—OH HO—C—H H—C—OH H—C—OH CH₂OH Gluconic acid & Isomers	CO₂H H—C—OH HO—C—H H—C—OH H—C—OH CO₂H Glucaric acid & Isomers	CO₂H H—C—H H—C—OH H—C—OH H—C—OH CH₂OH 2-Deoxyhexonic acids	CO₂H H—C—OH H—C—H H—C—OH H—C—OH CH₂OH 3-Deoxyhexonic acid		

Figure 1 Polyols identified in the Murchison and Murray carbonaceous meteorites. Some of the acids were also identified in their cyclic (lactone) form. When commercial standards were not available for mass spectral comparison, 'isomers' were identified by comparison of their mass spectrum and GC retention time to that of the listed member of the group. Indicated concentrations are per gram of meteorite. 1% for erythritol and threitol refers to their detector response versus that of glycerol. The detector response (DR) of 2-methyl glyceric acid (TMS derivative) was approximately one-sixth that of glyceric acid. The DR of ribonic acid (TMS) was roughly one-third that of threonic acid. In

the largest Murchison sample, the DRs of the 4C deoxy sugars were roughly 4–5 times larger than those of the 4C mono-acids (erythronic and threonic). The DRs of the 2-deoxypentonic acids were roughly 20 times less than that of ribonic acid. The DR of the 3-deoxyhexonic acid was approximately equal to that of the 2-deoxypentonic acids. The DR of the 3-deoxypentonic acid(s) was clearly less than that of the 3-deoxyhexonic acid. Asterisk indicates that 6C sugar monomers were not seen, but may be present in bound forms. Identifications of 2-deoxyhexonic acids and deoxydicarboxylic acids are tentative (see text).

decades before these studies (Murray was among the exceptions), thus increasing the chances of microbial contamination. A later reviewer concluded that a portion of the identified compounds could have been contaminants⁷. However, if contamination in the above studies had been severe, other sugars, including ribose, should have also been prominent. The authors stated that no other sugars were seen⁶. Although not part of the analyses, the simultaneous identification of sugar derivatives (decomposition products, and so on) would have been supporting evidence for—at least—aqueous sugar chemistry on a meteorite parent body (see below).

Our analyses of several Murchison extracts, and a Murray extract, show a variety of water-soluble polyols (Fig. 1). The compounds

that we identified include a sugar, dihydroxyacetone (sugars by definition are polyhydroxy aldehydes or ketones); sugar alcohols; sugar mono-acids; sugar di-acids; and deoxysugar acids (or 'saccharinic' acids). In general, the compounds follow the abiotic synthesis pattern of other meteorite classes of organic compounds¹: a general decrease in abundance with increasing carbon number within a class of compounds, and many, if not all, possible isomers present at a given carbon number. We positively identified many of the compounds shown (Fig. 2) by comparison of their mass spectra to commercially available standards (standards of deoxysugar acids were synthesized; see below). Murchison and Murray are similar with respect to the presence of individual polyols.

We found dihydroxyacetone in all Murchison extracts and the

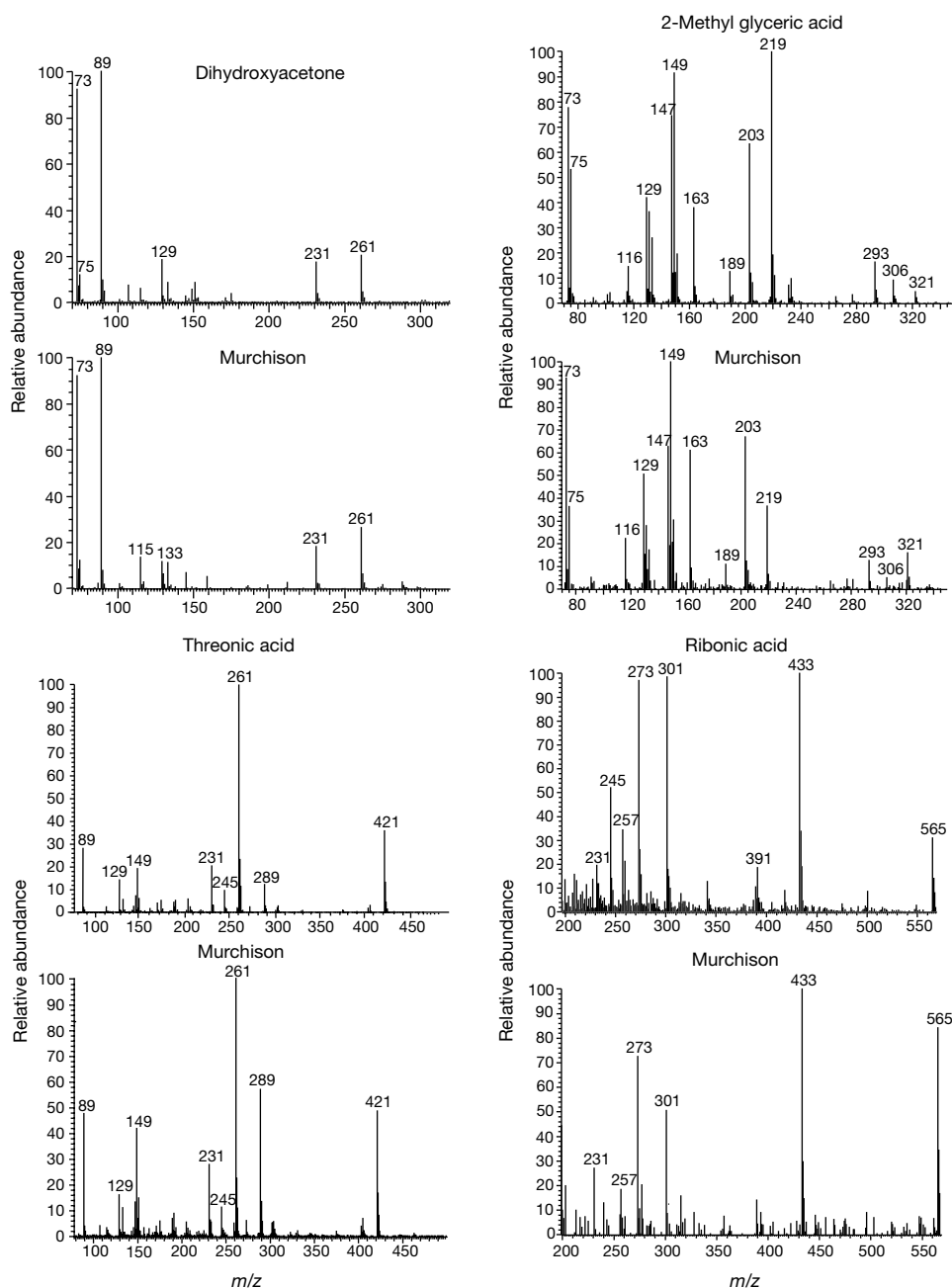


Figure 2 Selected mass spectra of *t*-BDMS and TMS derivatives of standards and corresponding compounds from Murchison and Murray. Dihydroxyacetone was characterized as the di-*t*-BDMS derivative; 2-methyl glyceric acid as the tri-TMS derivative; and threonic acid and ribonic acid as the tri-*t*-BDMS derivatives. *t*-BDMS

derivatives give a characteristic molecular ion minus 57 a.m.u. peak ($M^+ - 57$): for example, dihydroxyacetone(*t*-BDMS)₂ = 318 a.m.u., therefore $M^+ - 57 = 261$. Mass spectra of TMS-sugar derivatives have been described in detail¹³.

Murray extract. It was relatively abundant in one Murchison sample (Fig. 3). In two of the Murchison extracts and in the Murray extract the mass spectrum was weak (due to lower abundance), although it could be identified by ion searches for characteristic mass fragments (Fig. 2). If dihydroxyacetone was present because of contamination by microbes, a wider array of sugars should also have been seen^{8,9}. Results of analysis for the two-carbon (2C) analogue of dihydroxyacetone, glycolaldehyde, were inconclusive, and further analyses are in progress. Although we did not see 6C monomer sugars, we cannot rule out the presence of compounds containing them as there was suggestive evidence (mass spectra) of their presence in di-compounds in the largest Murchison sample.

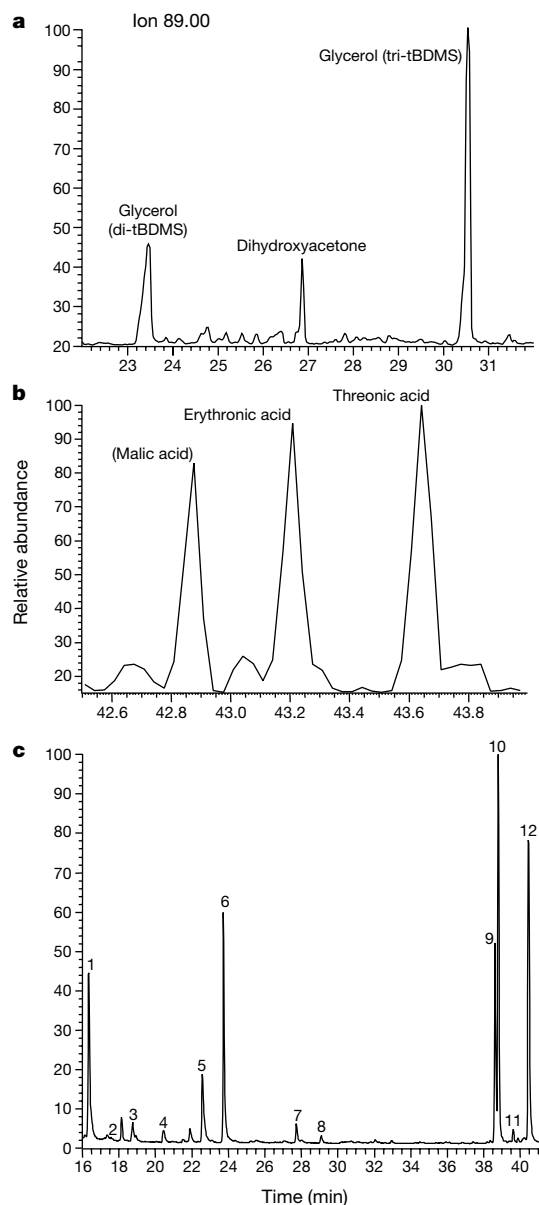


Figure 3 Polyols from meteorites and laboratory synthesis. **a**, Dihydroxyacetone and glycerol from a segment of a GC–MS total ion chromatogram of neutral compounds (water soluble) from Murchison. **b**, A portion of the total ion chromatogram of acidic compounds (*t*-BDMS derivatives) from Murchison. Malic acid and other mono-hydroxy acids are known constituents of Murchison¹. **c**, Reaction products (TMS derivatives) from alkaline reactions of glucose. Peak 1, glycerol; 2, ethylene glycol; 3, 2-methyl glyceric acid; 4, glyceric acid; 5, succinic acid; 6, 2,4-dihydroxybutyric acid; 7, threonic acid; 8, erythronic acid; 9, 10, 3-deoxyhexonic acids (6C); 11, 12, lactones of 3-deoxyhexonic acids.

The sugar alcohol series begins with glycerol; however, the 2C homologue, ethylene glycol, is also present and was seen in both meteorites. The series extends to at least 6C members. We initially identified these compounds as their *t*-BDMS derivatives, based on comparison of their mass spectra and retention times to standards. Ethylene glycol and glycerol are the most abundant of all identified polyols. The detector response of ethylene glycol (*t*-BDMS)₂ was roughly twice that of glycerol in the Murray sample and some of the Murchison samples. Glycerol is present in Murchison at approximately double the highest concentration found for an individual meteoritic amino acid¹ (Fig. 1). There is a significant drop in abundance from glycerol to higher alcohols. Excluding glycerol and ethylene glycol, the sugar alcohols are the lowest in abundance of all polyols. Concentrations of indicated polyols were determined by GC–MS and should be considered approximate, as there is significant variation with sample.

Glycerol, as well as some of the other sugar alcohols, is widespread in nature, and therefore a portion of the compound could be a contaminant either from analytical procedures (although none was seen in blanks) or from microbial action within the meteorite sample. However, given the relatively large amounts seen in all samples and the fact that the sugar alcohols are present in an abiotic distribution pattern, it is likely that a significant amount of the identified glycerol is indigenous to the meteorites. In this regard, the four-carbon (4C) alcohol, erythritol, which is found in nature, and its rare isomer¹⁰, threitol, are nearly equal in abundance in both Murchison and Murray. We did not see the three-carbon (3C) deoxysugar alcohols, 2,3-propanediol and 1,3-propanediol, in any sample. This may be important from a synthetic point of view (below). Also identified in Murchison, as their TMS derivatives, were inositols (6C cyclic alcohols). However, they need to be fully characterized in future work as some of the nine possible stereoisomers are widely distributed in nature¹⁰. Our isotopic analysis of combined neutral polyols found values of $\delta^{13}\text{C} = -5.89\text{‰}$ and $\delta\text{D} = +119\text{‰}$ (see Methods), suggesting that most are indigenous to the meteorite.

Our analyses show that the sugar mono-acids (aldonic acids) begin with the 3C member, glyceric acid, and extend to at least the 6C isomers (Fig. 1). (The 2C analogue of glyceric acid, glycolic acid, and other mono-hydroxy acids, are well known constituents of carbonaceous meteorites¹.) We find that the abundance of glyceric acid ($\sim 80 \text{ nmol g}^{-1}$) is similar to that of the more abundant meteoritic amino acids. The pattern of sugar acids that we find is also abiotic: for example, 2-methyl glyceric acid (2-MGA), a relatively uncommon compound in nature, is also abundant (Fig. 1 legend). Erythronic and threonic acids were nearly equal in abundance in one Murchison extract (Fig. 3)—but the ratio of erythronic to threonic acid varied from approximately two to three in other samples. Tartaric acid is the first member of the sugar dicarboxylic acid series. As with erythritol/threitol, tartaric acid (racemic) and its rare terrestrial isomer, mesotartaric acid, are nearly equal in abundance.

Deoxysugar acids (Fig. 1) are polyhydroxylated organic acids in which one or more hydroxyl groups (bonded to C) are replaced by hydrogen. The deoxyacids in Fig. 1 are less common in nature than sugars, and laboratory standards were not available commercially; however, they are well known products of alkaline reactions of sugars^{11,12}. We synthesized deoxysugar acids by known methods of alkaline degradation of 6C sugars; we also synthesized 2-deoxy 5C and 6C acids (Fig. 1) by acid oxidation of the corresponding 2-deoxysugars (see Methods). Their TMS and *t*-BDMS mass spectra were then compared to compounds in both meteorites and (TMS derivatives) to literature spectra¹³. The 4C members are relatively abundant (Fig. 1).

There are two (excluding enantiomers) possible straight-chain 5C 2-deoxymono acids (2-deoxypentonic acids), and both are apparently present in small amounts in Murchison. 2-deoxyhexonic

acids as well as 5C and 6C deoxydicarboxylic acids may also be present, but because of low abundances and resulting lower-quality mass spectra we cannot yet confirm their presence. Compounds of matching GC retention times and apparently of the same molecular masses (*t*-BDMS derivatives) are seen. In Murchison, we identified a 3-deoxyhexonic acid (a 'metasaccharinic' acid) (Fig. 1). At least one 3-deoxypentonic acid is tentatively identified. Other, unidentified, compounds with TMS mass spectra similar to those of deoxysugar acids¹³ are also present. These could include branched deoxysugar acids.

There may have been multiple mechanisms in interstellar space and on the Murchison and Murray parent bodies that synthesized either individual members or groups of the present compounds. Photolysis of small molecules—such as CO, NH₃, H₂O and so on—under simulated interstellar conditions has been shown to form the low-molecular-mass polyols ethylene glycol, glycerol and glyceric acid^{14,15}. If most of the present compounds were derived from sugars, a mechanism capable of producing such a variety of polyols might involve the 'formose reaction'¹⁶. In laboratory formose reactions, formaldehyde (CH₂O), in aqueous solution at neutral to basic pH, condenses to produce a variety of hydroxylated compounds (predominantly sugars but also sugar alcohols) of carbon number up to at least seven. Even in the case of production of only the smallest sugars, dihydroxyacetone and its isomer, glyceraldehyde, a variety of other sugars could then be produced by condensation in the presence of minerals in aqueous solution¹⁷. This may also have implications for the pre-biotic Earth in the case of a limited number of sugars (Fig. 1).

Several lines of evidence suggest that the formose reaction would have been possible on the parent bodies of carbonaceous meteorites: aqueous alteration occurred on the parent bodies and/or precursor interstellar ices^{18,19}; water extracts of carbonaceous meteorites are commonly slightly basic (pH ≈ 7–9); formaldehyde is a relatively abundant and ubiquitous molecule in interstellar space and comets²⁰; several small volatile carbonyl compounds are constituents of Murchison²¹; and glycolaldehyde, a possible formaldehyde reaction product necessary for the formation of higher sugars in the formose reaction, has recently been identified in interstellar space²². In addition, recent research employing ¹³C NMR to study the 'macromolecular carbon' of Murchison has shown that as much as 5% of this carbon has hydroxy and/or ether bonds²³. If formaldehyde condensation was a source of meteoritic polyol monomers, then it may have also contributed to this solid phase.

Sugars readily undergo reactions in alkaline solution^{12,24}, including the formation of multiple sugar acids from an individual sugar: for example, lactic, glyceric and ribonic acid from ribose. The presence of deoxysugar acids in our extracts (Fig. 1) is also consistent with alkaline sugars chemistry. Such acids have been reported to constitute as much as 10–30% of the alkaline degradation products of sugars¹¹. Consistent with previous studies, we found 2,4-dihydroxybutyric and other acids to be major products of the alkaline degradation of glucose (Fig. 3). 3-deoxyhexonic acids have been reported to be relatively abundant in alkaline sugar experiments²⁵, but we find them to be relatively minor products in Murchison. The extent and conditions of laboratory reactions compared to alteration on the meteorite parent body, as well as the presence of formation mechanisms other than the above, could account for differences between our laboratory experiments and observations. □

Methods

Extraction and detection of compounds

In some cases, sample extraction, preparation procedures and GC conditions were similar to those used previously²⁶. Results of sugar analysis include samples extracted under helium (99.999%). Cation exchange procedures were as previously described²⁶. Anion exchange resins were either Biorad AG4-X4 (OH[−] form) or Biorad AG1-X8 (acetate form). Derivatization reagents were *N*-methyl-*N*-(*t*-butyldimethylsilyl)trifluoroacetamide with

1% *t*-butyldimethylsilyl chloride, Regis Chemical Co. (to obtain *t*-BDMS derivatives), and bis(trimethylsilyl) trifluoroacetamide (BSTFA) with 1% trimethylchlorosilane, Alltech Assoc. (TMS derivatives). Pyridine was used as the solvent in a ratio of 4/1 (pyridine/reagent) in most cases. In most GC–MS analyses, a Finnigan ion trap GCQ GC–MS was used. A DB-17 (30 or 60 m × 0.25 mm) fused-silica capillary column (J & W Scientific) was most often used for separations.

Synthesis of sugar derivatives

In alkaline degradation of sugars, individual sugars were heated (80–100 °C) in 1–8 M NaOH for between 5 min and 2 h. 2-Deoxyribonic acid and 2-deoxygluconic acid standards were obtained by HNO₃ oxidation of 2-deoxyribose and 2-deoxyglucose, respectively. Most acids were identified in their straight-chain form by bringing samples to slightly basic pH before derivatization procedures.

Isotope measurements of neutral compounds

Neutral compounds, including polyols and amides, were obtained by ion exchange chromatography of a Murchison extract as described previously²⁶. Polyols were isolated by acid hydrolysis (HCl) and re-subjecting the fraction to cation and anion exchange chromatography. The isolated polyols were then placed in a quartz tube and dried by rotary evaporation. Copper oxide was then placed in the tube and higher vacuum was used for more complete drying. The samples were then sealed and combusted to carbon dioxide and water at 800 °C. The carbon dioxide and water were collected with a procedure similar to that of ref. 27. The water was converted to hydrogen by the method of ref. 28 using a zinc catalyst (J. M. Hayes, Indiana University). The carbon dioxide and hydrogen were then analysed with 6-60-Nuclide and 3-60-HD Nuclide mass spectrometers. The heavy to light isotope ratio, H/L, for each element (X), reported as per mil (‰), is defined as $\delta X = \{[(H/L)_{\text{sample}}/(H/L)_{\text{standard}}] - 1\} \times 1,000$. The reported values (see text) are uncorrected for terrestrial carbon and hydrogen contributed during analytical procedures and thus should be considered lower limits. Even so, they lie outside the range of typical terrestrial organic compounds²⁹, and are consistent with values obtained for other classes of meteoritic organic compounds¹.

Received 14 August; accepted 30 October 2001.

1. Cronin, J. R. & Chang, S. in *Chemistry of Life's Origins* (eds Greenberg, J. M., Pirronello, V. & Mendoza-Gomez, C.) 209–258 (Kluwer, Dordrecht, 1993).
2. Oro, J. Comets and the formation of biochemical compounds on the primitive earth. *Nature* **190**, 389–390 (1961).
3. Anders, E. Pre-biotic organic matter from comets and asteroids. *Nature* **342**, 255–257 (1989).
4. Chyba, C. F. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125–132 (1992).
5. Degens, E. T. & Bajor, M. Amino acids and sugars in the Bruderheim and Murray meteorite. *Naturwissenschaften* **49**, 605–606 (1962).
6. Kaplan, I. R., Degens, E. T. & Reuter, J. H. Organic compounds in stony meteorites. *Geochim. Cosmochim. Acta* **27**, 805–834 (1963).
7. Hayes, J. M. Organic constituents of meteorites—a review. *Geochim. Cosmochim. Acta* **31**, 1395–1440 (1967).
8. Amelung, W., Cheshire, M. V. & Guggenberger, G. Determination of neutral and acidic sugars in soil by capillary gas-liquid chromatography after trifluoroacetic acid hydrolysis. *Soil Biol. Biochem.* **28**, 1631–1639 (1997).
9. Gilbert, J., Fox, A. & Morgan, S. L. Carbohydrate profiling of bacteria by gas chromatography-mass spectrometry: Chemical derivatization and analytical pyrolysis. *Eur. J. Clin. Microbiol.* **6**, 715–723 (1987).
10. Brimacombe, J. S. & Webber, J. M. in *The Carbohydrates, Chemistry and Biochemistry* (eds Pigman, W. & Horton, D.) 479–518 (Academic, New York, 1972).
11. Pigman, W. & Anet, E. F. L. J. in *The Carbohydrates, Chemistry and Biochemistry* (eds Pigman, W. & Horton, D.) 165–194 (Academic, New York, 1972).
12. De Bruijn, J. M., Kieboom, A. P. G. & Van Bekkum, H. Reactions of monosaccharides in aqueous alkaline solutions. *Sugar Tech. Rev.* **13**, 21–52 (1986).
13. Petersson, G. Mass spectrometry of aldonic and deoxyaldonic acids as trimethylsilyl derivatives. *Tetrahedron* **26**, 3413–3428 (1970).
14. Agarwal, V. K. et al. Photochemical reactions in interstellar grain photolysis of CO, NH₃, and H₂O. *Origins Life* **16**, 21–40 (1986).
15. McDonald, G. D. et al. Production and chemical analysis of cometary ice tholins. *Icarus* **122**, 107–117 (1996).
16. Langenbeck, W. Die formaldehydkondensation als organische autokatalyse. *Naturwissenschaften* **30**, 30–34 (1942).
17. Weber, A. L. Prebiotic sugars: hexose and hydroxy acid synthesis from glyceraldehyde catalyzed by iron (III) hydroxide oxide. *J. Mol. Evol.* **35**, 1–6 (1992).
18. Bunch, T. E. & Chang, S. Carbonaceous chondrites—II. Carbonaceous chondrites phyllosilicates and light element geochemistry as indicators of parent body processes and surface conditions. *Geochim. Cosmochim. Acta* **44**, 1543–1577 (1980).
19. Zolensky, M. & McSween, J. Y. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 114–143 (Univ. Arizona Press, Tucson, 1988).
20. Irvine, W. M. The composition of interstellar molecular clouds. *Space Sci. Rev.* **90**, 203–218 (1999).
21. Jungclaus, G. A., Yuen, G. U., Moore, C. B. & Lawless, J. G. Evidence for the presence of low molecular weight alcohols and carbonyl compounds in the Murchison meteorite. *Meteoritics* **11**, 231–237 (1976).
22. Hollis, J. M., Lovas, F. J. & Jewell, P. R. Interstellar glycolaldehyde: the first sugar. *Astrophys. J.* **540**, L107–L110 (2000).
23. Cody, G. D., Alexander, C. M. O'D. & Tera, F. New insights into the chemistry of Murchison organic macromolecule using high field ¹³C solid state NMR. *Lunar Planet. Sci. Conf.* **XXX**, 1582–1583 (1999).
24. Larraalde, R., Robertson, M. P. & Miller, S. L. Rates of decomposition of ribose and other sugars: Implications for chemical evolution. *Proc. Natl Acad. Sci. USA* **92**, 8158–8160 (1995).

25. Lowendahl, L., Petersson, G. & Samuelson, O. Formation of carboxylic acids by degradation of carbohydrates during kraft cooking of pine. *Technic. Assoc. Pulp Paper Ind.* **59**, 118–120 (1976).
26. Cooper, G. W. & Cronin, J. R. Linear and cyclic aliphatic carboxamides of the Murchison meteorite: Hydrolyzable derivatives of amino acids and other carboxylic acids. *Geochim. Cosmochim. Acta* **59**, 1003–1015 (1995).
27. Epstein, S., Krishnamurthy, R. V., Cronin, J. R., Pizzarello, S. & Yuen, G. U. Unusual stable isotope ratios in amino acid and carboxylic acid extracts from the Murchison meteorite. *Nature* **326**, 477–479 (1987).
28. Coleman, M. L. & Moore, M. P. Direct reduction of sulfates to sulfur dioxide for isotopic analysis. *Anal. Chem.* **50**, 1594–1595 (1978).
29. Des Marais, D. J. Isotopic evolution of the biogeochemical carbon cycle during the Proterozoic eon. *Org. Geochem.* **27**, 185–193 (1997).

Acknowledgements

We thank A. Weber and J. Cronin for discussions and comments on the manuscript; and C. Asiyo and T. Esposito for assistance with figures. This work was supported by the Exobiology Program of NASA.

Correspondence and requests for materials should be addressed to G.C.
(e-mail: gcooper@mail.arc.nasa.gov).

Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance

Lieven M. K. Vandersypen^{*†}, Matthias Steffen^{*†}, Gregory Breyta^{*}, Costantino S. Yannoni^{*}, Mark H. Sherwood^{*} & Isaac L. Chuang^{*†}

^{*} IBM Almaden Research Center, San Jose, California 95120, USA

[†] Solid State and Photonics Laboratory, Stanford University, Stanford, California 94305-4075, USA

The number of steps any classical computer requires in order to find the prime factors of an l -digit integer N increases exponentially with l , at least using algorithms known at present¹. Factoring large integers is therefore conjectured to be intractable classically, an observation underlying the security of widely used cryptographic codes^{1,2}. Quantum computers³, however, could factor integers in only polynomial time, using Shor's quantum factoring algorithm^{4–6}. Although important for the study of quantum computers⁷, experimental demonstration of this algorithm has proved elusive^{8–10}. Here we report an implementation of the simplest instance of Shor's algorithm: factorization of $N = 15$ (whose prime factors are 3 and 5). We use seven spin-1/2 nuclei in a molecule as quantum bits^{11,12}, which can be manipulated with room temperature liquid-state nuclear magnetic resonance techniques. This method of using nuclei to store quantum information is in principle scalable to systems containing many quantum bits¹³, but such scalability is not implied by the present work. The significance of our work lies in the demonstration of experimental and theoretical techniques for precise control and modelling of complex quantum computers. In particular, we present a simple, parameter-free but predictive model of decoherence effects¹⁴ in our system.

Shor's factoring algorithm works by using a quantum computer to quickly determine the period of the function $f(x) = a^x \bmod N$ (the remainder of a^x divided by N), where a is a randomly chosen small number with no factors in common with N ; from this period, number-theoretic techniques can be used to factor N with high probability^{4–6}. The two main components of the algorithm, modular exponentiation (computation of $a^x \bmod N$) and the inverse quantum Fourier transform (QFT) take only $O(l^3)$ operations^{4–6}. Classically, in contrast, prime factorization takes $O(2^{l/3})$ operations¹, which quickly becomes intractable as l increases.

The simplest meaningful instance of Shor's algorithm is factorization of $N = 15$ (ref. 7)—the algorithm fails for N even or a prime power. Even for such a small N , quantum factorization poses at present a significant experimental challenge: it requires coherent control over seven quantum bits (qubits) in the course of a long sequence of controlled interactions, even after maximal reduction of the quantum circuit; including the state initialization, interactions between almost all pairs of qubits are needed. In comparison with earlier work^{8–10}, this experiment thus puts extremely high demands on the spin–spin coupling network, the degree of control over the hamiltonian and the spin coherence times. Furthermore, numerically predicting the outcome of these experiments has been considered impractical owing to the enormous size of the state space transformations, which are described by $\sim 4^7 \times 4^7$ real parameters if decoherence effects are included.

Implementation of the algorithm can be broken into four distinct steps (Fig. 1a), with the most complex being the computation of $f(x) = a^x \bmod N$ for 2^n values of x in parallel. Following standard classical circuit techniques, this is performed by utilizing the identity $a^x = a^{2^{n-1}x_{n-1} + \dots + 2^1x_1 + x_0}$, where x_k are the binary digits of x . Modular exponentiation thus consists of serial multiplication by $a^{2^k} \bmod N$ for all k ($0 \leq k \leq n-1$) for which $|x_k\rangle = |1\rangle$. The powers a^{2^k} can be efficiently pre-computed on a classical machine by repeated squaring of a . For $N = 15$, a may be 2, 4, 7, 8, 11, 13 or 14. If we happen to pick $a = 2, 7, 8$ or 13 , we find that $a^4 \bmod 15 = 1$, and therefore all $a^{2^k} \bmod N = 1$ for $k \geq 2$. In this case, $f(x)$ simplifies to multiplications controlled by just two bits, x_0 and x_1 . If $a = 4, 11$ or 14 , then $a^2 \bmod 15 = 1$, so only x_0 is relevant. Thus, the first register can be as small as two qubits ($n = 2$); however, three qubits ($n = 3$) allow for the possibility of detecting more periods, and thus constitutes a more stringent test of the modular exponentiation and QFT (M.S. *et al.*, manuscript in preparation). Together with the $m = \lfloor \log_2 15 \rfloor = 4$ qubits to hold $f(x)$, we need seven qubits in total (Fig. 1b). We implemented this algorithm and tested it on two representative parameter choices: $a = 11$ (an 'easy' case) and $a = 7$ (a 'difficult' case).

The custom-synthesized molecule used as the quantum computer for this experiment contains five ^{19}F and two ^{13}C spin-1/2 nuclei as qubits (Fig. 2). In a static magnetic field, each spin i has two discrete energy eigenstates, $|0\rangle$ (spin-up) and $|1\rangle$ (spin-down), described by the hamiltonian $H_0 = -\sum_i \hbar \omega_i I_{zi}$, where $\omega_i/2\pi$ is the transition frequency between $|0\rangle$ and $|1\rangle$ and I_z is the $\hat{z}$ component of the spin angular momentum operator. All seven spins in this molecule are remarkably well separated in frequency $\omega_i/2\pi$, and interact pairwise via the J -coupling, described by $H_J = \sum_{i < j} 2\pi \hbar J_{ij} I_{zi} I_{zj}$ (ref. 15).

The desired initial state of the seven qubits is $|\psi_i\rangle = |0000001\rangle$ (Fig. 1). However, experimentally we start from thermal equilibrium. The density matrix is then given by $\rho_{\text{th}} = e^{-H_0/k_B T}/2^7$, with $k_B T \gg \hbar \omega_i$ at room temperature so each spin is in a statistical mixture of $|0\rangle$ and $|1\rangle$ (Fig. 3a). We converted ρ_{th} into a 7-spin effective pure state^{11,12} ρ_1 via temporal averaging⁹ (step 0); ρ_1 constitutes a suitable initial state for Shor's factoring algorithm because it generates the same signal as $|\psi_i\rangle$ (Fig. 3b), up to a proportionality constant^{11,12}. Although ρ_1 is highly mixed and in fact remains separable under unitary transforms, the observed dynamics under multiple qubit operations such as in Shor's algorithm apparently remain hard to simulate classically^{16–18}.

The quantum circuit of Fig. 1 was realized with a sequence of ~ 300 ($a = 7$) spin-selective radio-frequency (r.f.) pulses separated by time intervals of free evolution under the hamiltonian (Fig. 4). The pulse sequence is designed such that the resulting transformations of the spin states correspond to the computational steps in the algorithm. Upon completion of this sequence, we estimate the state of the first three qubits, $\rho \sim \sum_i w_i c^3/r |c^3/r\rangle$, via nuclear magnetic resonance (NMR) spectroscopy. In the experiment, an ensemble of independent quantum computers rather than a single quantum

computer was used, so the measurement gives the bit-wise average value of $8c/r$, instead of a sample of $8c/r$. This is sufficient to determine r in the present experiment, but for larger N a continued fractions algorithm will need to be performed on the quantum computer¹¹, requiring additional qubits. From r , at least one factor of N is given by the greatest common denominator (g.c.d.) of $a^{r/2} \pm 1$ and N (with probability greater than $1/2$); the g.c.d. can be computed efficiently using Euclid's algorithm on a classical computer².

The experimental spectra acquired upon completion of the easy case ($a = 11$) of Shor's algorithm (Fig. 3c) clearly indicate that qubits 1 and 2 are in $|0\rangle$ (spectral lines up), and that 3 is in an equal mixture of $|0\rangle$ and $|1\rangle$ (lines up and down, and the integral of the spectrum equal to zero). With qubit 3 the most significant qubit after the inverse QFT¹⁹, the first register is thus in a mixture of $|000\rangle$ and $|100\rangle$, or $|0\rangle$ and $|4\rangle$ in decimal notation. The periodicity in the amplitude of $|y\rangle$ is thus 4, so $r = 2^n/4 = 2$ and we find that g.c.d.(11^{2/2} $\pm$ 1, 15) = 3, 5. The prime factors thus unambiguously derive from the output spectra.

From analogous spectra for the difficult case ($a = 7$; Fig. 3d), we see that qubit 1 is in $|0\rangle$, and qubits 2 and 3 are in a mixture of $|0\rangle$ and $|1\rangle$. The register is thus in a mixture of $|000\rangle$, $|010\rangle$, $|100\rangle$ and $|110\rangle$, or $|0\rangle$, $|2\rangle$, $|4\rangle$ and $|6\rangle$. The periodicity in the amplitude of $|y\rangle$ is now 2, so $r = 8/2 = 4$ and g.c.d.(7^{4/2} $\pm$ 1, 15) = 3, 5. Thus, even after the long and complex pulse sequence of the difficult case (Fig. 4), the experimental data conclusively indicate the successful execution of Shor's algorithm to factor 15.

Nevertheless, there are obvious discrepancies between the measured and ideal spectra, most notably for the difficult case. Using a numerical model, we have investigated whether these deviations

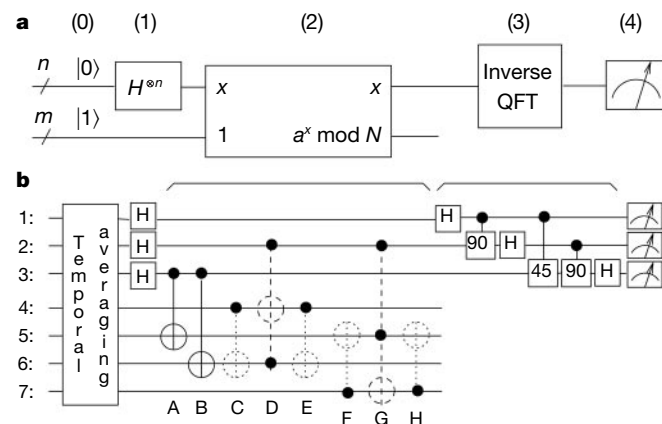


Figure 1 Quantum circuit for Shor's algorithm. **a**, Outline of the quantum circuit. Wires represent qubits, and boxes represent operations. Time goes from left to right. (0) Initialize a first register of $n = 2\lceil \log_2 N \rceil$ qubits to $|0\rangle \otimes \dots \otimes |0\rangle$ (for short $|0\rangle$) and a second register of $m = \lceil \log_2 N \rceil$ qubits to $|0\rangle \otimes \dots \otimes |0\rangle \otimes |1\rangle$. (1) Apply a Hadamard transform H to the first n qubits, so the first register reaches $\sum_{x=0}^{2^n-1} |x\rangle / \sqrt{2^n}$. (2) Multiply the second register by $f(x) = a^x \text{ mod } N$ (for some random $a < N$ which has no common factors with N), to get $|\psi_2\rangle = \sum_{x=0}^{2^n-1} |x\rangle \otimes |a^x \text{ mod } N\rangle / \sqrt{2^n}$. As the first register is in a superposition of 2^n terms $|x\rangle$, the modular exponentiation is computed for 2^n values of x in parallel. (3) Perform the inverse QFT on the first register¹⁹, giving $|\psi_3\rangle = \sum_{y=0}^{2^n-1} \sum_{x=0}^{2^n-1} e^{2\pi i xy/2^n} |y\rangle |a^x \text{ mod } N\rangle / 2^n$, where interference causes only terms $|y\rangle$ with $y = c2^n/r$ (for integer c) to have a substantial amplitude, with r the period of $f(x)$. (4) Measure the qubits in the first register. On an ideal single quantum computer, the measurement outcome is $c2^n/r$ for some c with high probability, and r can be quickly deduced from $c2^n/r$ on a classical computer via continued fractions². **b**, Detailed quantum circuit for the case $N = 15$ and $a = 7$. Control qubits are marked by filled circles; $\oplus$ represents a NOT operation and 90 and 45 represent $\hat{z}$ rotations over these angles. The gates shown in dotted lines can be removed by optimization, and the gates shown in dashed lines can be replaced by simpler gates (see Methods).

could be caused by decoherence. A full description of relaxation for the seven coupled spins involves almost $4^7 \times 4^7$ degrees of freedom, and requires knowledge of physical properties of the molecule which are not available^{20,21}. In order to get a first estimate of the effect of decoherence during the factoring pulse sequence, we assume that each spin experiences independent stochastic relaxation with correlation timescales $\ll 1/\omega_i$. This permits the use of the phenomenological Bloch equations²², with just two time constants per spin (T_1 and T_2). We implemented this decoherence model for seven coupled spins via the operator sum representation²³ $\rho \mapsto \sum_k E_k \rho E_k^\dagger$ ($\sum_k E_k^\dagger E_k = I$), starting from existing single spin models²⁴ of generalized amplitude damping (GAD, T_1),

$$E_0 = \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \quad E_1 = \sqrt{p} \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix} \quad (1)$$

$$E_2 = \sqrt{1-p} \begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{bmatrix}, \quad E_3 = \sqrt{1-p} \begin{bmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{bmatrix}$$

and phase damping (PD, related to T_2),

$$E_0 = \sqrt{\lambda} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad E_1 = \sqrt{1-\lambda} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (2)$$

with $\gamma = 1 - e^{-t/T_1}$, $p = 1/2 + \hbar\omega/4k_B T$ and $\lambda \sim (1 + e^{-t/T_2})/2$. The following observations simplify the extension of these separate single spin descriptions to an integrated model for seven spins: (1) GAD (and PD) error operators acting on different spins commute; (2) the E_k for GAD commute with the E_k for PD when applied to arbitrary ρ ; and (3) PD commutes with the ideal unitary

i	$\omega_i/2\pi$	$T_{1,i}$	$T_{2,i}$	J_{7i}	J_{6i}	J_{5i}	J_{4i}	J_{3i}	J_{2i}
1	-22052.0	5.0	1.3	-221.0	37.7	6.6	-114.3	14.5	25.16
2	489.5	13.7	1.8	18.6	-3.9	2.5	79.9	3.9	
3	25088.3	3.0	2.5	1.0	-13.5	41.6	12.9		
4	-4918.7	10.0	1.7	54.1	-5.7	2.1			
5	15186.6	2.8	1.8	19.4	59.5				
6	-4519.1	45.4	2.0	68.9					
7	4244.3	31.6	2.0						

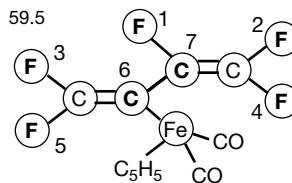


Figure 2 Structure and properties of the quantum computer molecule, a perfluorobutadienyl iron complex with the inner two carbons ^{13}C -labelled. Based on the measured $J_{13,19\text{F}}$ values, we concluded that the placement of the iron is as shown, different from that derived in ref. 29 from infrared spectroscopy. The table gives the $\omega_i/2\pi$ (in Hz) at 11.7 T, relative to a reference frequency of ~ 470 MHz and ~ 125 MHz for ^{19}F and ^{13}C respectively, the longitudinal (T_1 , inversion recovery) and transverse (T_2 , estimated from a single spin-echo sequence) relaxation time constants (in s), and the J -couplings (in Hz). Ethyl (2- ^{13}C)bromoacetate (Cambridge Isotope Laboratories, Inc.) was converted to ethyl 2-fluoroacetate by heating with AgF, followed by hydrolysis to sodium fluoroacetate using NaOH in MeOH. This salt was converted to 1,1,1,2-tetrafluoroethane using MoF_6 , and was subsequently treated with two equivalents of n -butyl lithium followed by I_2 to provide trifluoroiodoethene. Half of the ethene was converted to the zinc salt, which was recombined with the remaining ethene and coupled using $\text{Pd}(\text{Ph}_3\text{P})_4$ to give (2,3- ^{13}C)hexafluorobutadiene. The end product was obtained by reacting this butadiene with the anion obtained from treating $[(\pi\text{-C}_5\text{H}_5)\text{Fe}(\text{CO})_2]_2$ with sodium amalgam²⁹. The product was purified with column chromatography, giving a total yield of about 5%. The sample, at 0.88 ± 0.04 mol% in perdeuterated diethyl ether was dried using 3-Å molecular sieves, filtered through a 0.45- μm syringe filter, and flame-sealed in the NMR sample tube using three freeze-thaw vacuum degassing cycles. All experiments were performed at 30 °C.

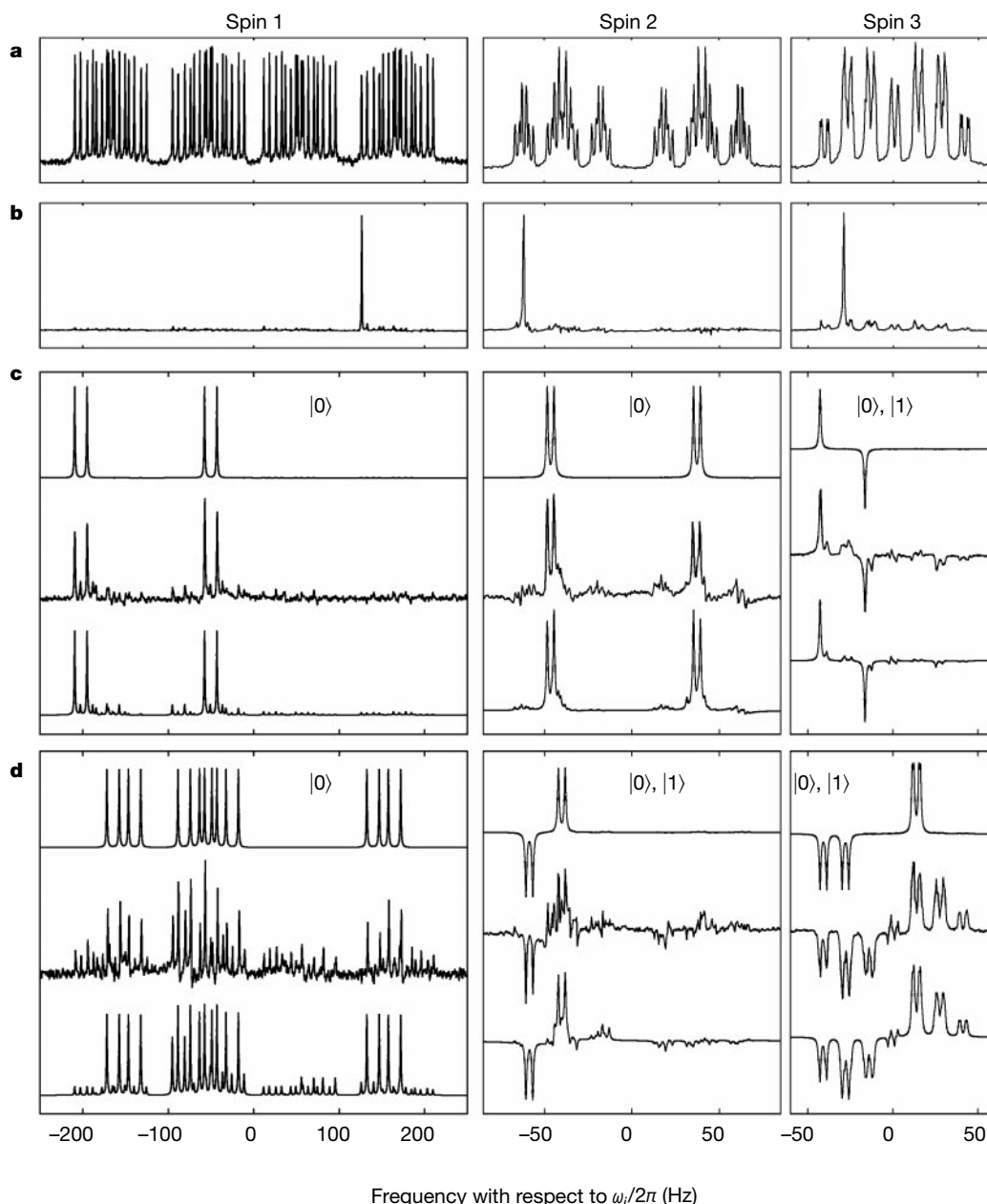


Figure 3 NMR spectra at different stages in the computation. **a**, Experimentally measured thermal equilibrium spectra (real part), acquired after a read-out pulse on spin i has tipped the spin from $|0\rangle$ ($+\hat{z}$) or $|1\rangle$ ($-\hat{z}$) into the $\hat{x}$ – $\hat{y}$ plane, where it induces a voltage oscillating at $\omega_i/2\pi + \sum_j \pm J_{ij}/2$ (where the sign depends on the state of the other spins) in a transverse r.f. coil placed near the sample. This voltage was recorded by a phase-sensitive detector and Fourier transformed to obtain a spectrum, with the phase set such that positive (negative) lines correspond to a spin in $|0\rangle$ ($|1\rangle$) before the readout pulse. Frequencies are in hertz, and with respect to $\omega/2\pi$. **b**, Experimental spectra for the

effective pure ground state. As desired, only one line is retained in each multiplet with its position depending on strength and sign of the J -couplings. Here, the transition corresponds to all other spins in $|0\rangle$. The state ρ_i is obtained from this state by applying a NOT on spin 7. **c**, Output spectra of the easy case of Shor's algorithm ($a = 11$). The top traces are the ideally expected spectra, the middle traces are the experimental data, and the bottom traces are simulations which incorporate decoherence effects (see text). Each trace was rescaled separately. **d**, Similar set of spectra as in **c**, but for the difficult case ($a = 7$).

time evolution e^{-iHt} ($H = H_0 + H_J$). However, GAD does not commute with e^{-iHt} , and PD and GAD do not commute with the ideal unitary evolution during r.f. pulses. Nevertheless, we have treated these as commuting transformations, such that all of the processes which occur simultaneously can be modelled sequentially.

Specifically, the model simulates a delay time of duration t_d by e^{-iH_d} followed by GAD acting on spin 1 for a duration t_d , then GAD acting on spin 2 and so forth, followed by PD acting serially on each spin. Similarly, a shaped pulse of duration t_p was modelled by a delay time of duration t_p ($e^{-iH_0 t_p}$, GAD and PD) followed by an instantaneous pulse. Using this simple model, the 7-spin simulation of the

complete Shor pulse sequence, including 36 temporal averaging sequences, required only a few minutes on an IBM quad POWER3-II processor machine. Measured values of T_1 and T_2 (Fig. 2) were used in the model.

The output spectra predicted by this parameter-free decoherence model are also shown in Fig. 3c and d. Although some discrepancies between the data and the simulations remain (due to the approximations in the model, as well as due to experimental imperfections such as r.f. inhomogeneity, imperfect calibrations, incomplete field drift compensation and incomplete unwinding of coupled evolution during the r.f. pulses; M.S. *et al.*, manuscript in preparation),

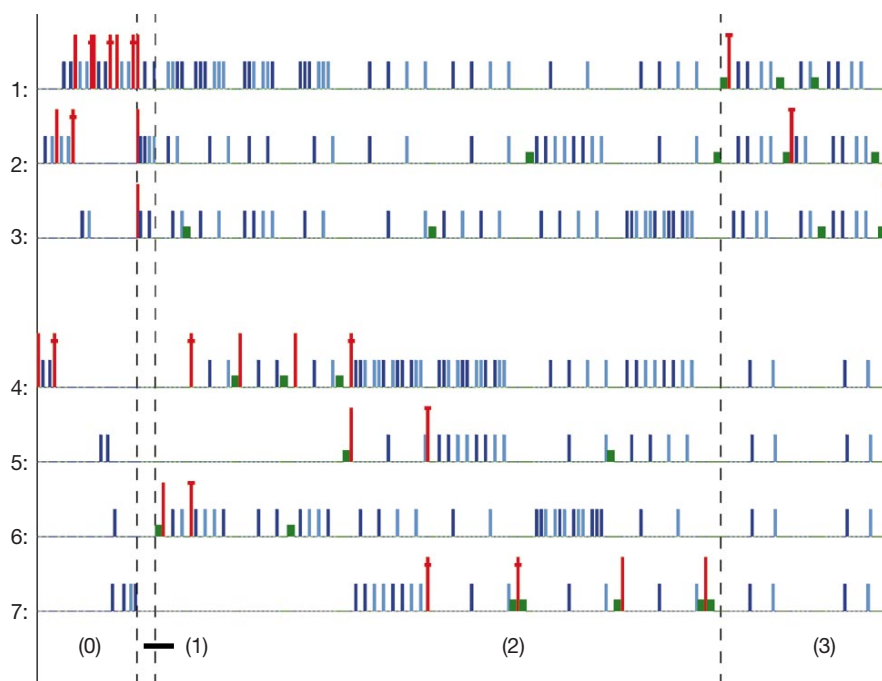


Figure 4 Pulse sequence for implementation of the quantum circuit of Fig. 1 for $a = 7$. The tall red lines represent 90° pulses selectively acting on one of the seven qubits (horizontal lines) about positive $\hat{x}$ (no cross), negative $\hat{x}$ (lower cross) and positive $\hat{y}$ (top cross). Note how single 90° pulses correspond to Hadamard gates, and pairs of such pulses separated by delay times correspond to two-qubit gates. The smaller blue lines

denote 180° selective pulses used for refocusing³⁰ about positive (darker shade) and negative $\hat{x}$ (lighter shade). Rotations about $\hat{z}$ are denoted by smaller and thicker green rectangles, and were implemented with frame-rotations. Time delays are not drawn to scale. The vertical dashed black lines visually separate the steps of the algorithm; step (0) shows one of the 36 temporal averaging sequences.

the model agrees well with the large non-idealities of the data. The predictive value of the model was further confirmed via independent test experiments.

This is, to our knowledge, the first NMR quantum computation experiment for which decoherence is the dominant source of errors⁸; the demands of Shor's algorithm clearly push the limits of the current molecule, despite its exceptional properties. At the same time, the good agreement between the measured and simulated spectra suggests that the degree of unitary control in the experiment was very high, which bodes well for related proposed implementations of quantum computers^{25,26}. Finally, we note that our parameter-free decoherence model, a predictive tool for modelling quantum errors in this complex system, provides an avenue for future design simulation of quantum computers. □

Methods

Experiments were performed at the IBM Almaden Research Center with an 11.7-T (500-MHz) Oxford Instruments magnet, a custom-modified four-channel Varian Unity INOVA spectrometer, and a Nalorac HFX probe. We extended the techniques of ref. 9 for serving multiple nuclei per channel, for reducing cross-talk between r.f. pulses on different spins and for sending simultaneous pulses. We used spin-selective Hermite-180 and Gaussian-90 pulses¹⁵, shaped in 256 steps, with a duration of 0.22 to ~ 2 ms. A technique to compensate for coupling effects during the selective pulses was developed and implemented via 'negative delay' times before and after the pulse. The amount of negative evolution needed depends on the pulse shape, and was pre-computed via numerical simulations.

To create an effective pure ground state of all seven spins, which has (to our knowledge) not been done before, we used a two-stage extension of the scheme of ref. 9, necessary because ω_{13C} is very different from ω_{19F} . The five ^{19}F spins are made effective pure via the summation of nine experiments, each with a different sequence of CN_{ij} and N_i gates (CN_{ij} stands for a controlled-NOT operation, which flips the target qubit j if and only if the control i is in $|1\rangle$; N_i simply flips i)²⁴. These nine experiments are executed four times, each time with different additional CN_{ij} and N_p such that the two ^{13}C spins are also made effective pure. Summation of the $4 \times 9 = 36$ experiments along with a NOT on spin 7 gives ρ_1 . The state preparation sequences were designed to be as short as possible (~ 200 ms) by making optimal use of the available coupling network.

Multiplication of $y = 1$ by $a \bmod 15$ controlled by x_0 (qubit 3) was replaced by

controlled-addition of $(a - 1) \bmod 15$. For $a = 11$, this is done by $CN_{34}CN_{36}$ and for $a = 7$ by $CN_{35}CN_{36}$ (gates A and B of Fig. 1b). Multiplication of y by $7^2 \bmod 15$ is equivalent to multiplication of y by $4 \bmod 15$, which reduces to swapping y_0 with y_2 and y_1 with y_3 . Both swap operations must be controlled by x_1 , which can be achieved²⁷ via gates C, D, E and F, G, H of Fig. 1b. This quantum circuit can be further simplified by a quantum analogue to peephole compiler optimization²⁸, which we consider should become standard in future quantum compilers: (1) the control qubit of gate C is $|0\rangle$, so C was suppressed; (2) similarly, F was replaced by N_5 ; (3) gates H and E are inconsequential for the final state of the first register, so they were omitted; (4) the targets of the doubly controlled NOT gates D and G are in a basis state, so they can be implemented as $CY_{34}CY_{24}CY_{23}$ and $CY_{27}CY_{27}CY_{27}$ (CZ_{ij} stands for a $90^\circ \hat{z}$ rotation of j if and only if i is in $|1\rangle$); (5) the refocusing schemes were kept as simple as possible. To this end, A was carried out after E. We did refocus inhomogeneous dephasing for all spins in the transverse plane. Residual couplings with the cyclopentadienyl protons were decoupled by continuous on-resonance low-power irradiation using a separate power amplifier and additional power combiners and r.f. filters. After all these simplifications, the pulse sequence for $7^x \bmod 15$ was ~ 400 ms long. The inverse QFT was implemented as shown in Fig. 1b and took ~ 120 ms. The duration of the complete sequence for the Shor algorithm was thus up to ~ 720 ms. A detailed report on these methods will be published elsewhere (M.S. *et al.*, manuscript in preparation).

Received 13 July; accepted 8 October 2001.

- Knuth, D. E. *The Art of Computer Programming* Vol. 2, *Seminumerical Algorithms* (Addison-Wesley, Reading, Massachusetts, 1998).
- Koblitz, N. *A Course in Number Theory and Cryptography* (Springer, New York, 1994).
- Bennett, C. H. & DiVincenzo, D. P. Quantum information and computation. *Nature* **404**, 247–255 (2000).
- Shor, P. in *Proc. 35th Annu. Symp. on the Foundations of Computer Science* (ed. Goldwasser, S.) 124–134 (IEEE Computer Society Press, Los Alamitos, California, 1994).
- Shor, P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509 (1997).
- Ekert, A. & Jozsa, R. Quantum computation and Shor's factoring algorithm. *Rev. Mod. Phys.* **68**(3), 733–753 (1996).
- Beckman, D., Chari, A. N., Devabhaktuni, S. & Preskill, J. Efficient networks for quantum factoring. *Phys. Rev. A* **54**, 1034–1063 (1996).
- Jones, J. A. NMR quantum computation. *Prog. NMR Spectrosc.* **38**, 325–360 (2001).
- Vandersypen, L. M. K. *et al.* Experimental realization of an order-finding algorithm with an NMR quantum computer. *Phys. Rev. Lett.* **85**, 5452–5455 (2000).
- Knill, E., Laflamme, R., Martinez, R. & Tseng, C.-H. An algorithmic benchmark for quantum information processing. *Nature* **404**, 368–370 (2000).
- Gershenfeld, N. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).

12. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci.* **94**, 1634–1639 (1997).
13. Schulman, L. & Vazirani, U. in *Proc. 31st ACM Symp. on Theory of Computing* 322–329 (Association for Computing Machinery, New York, 1999).
14. Chuang, I. L., Laflamme, R., Shor, P. & Zurek, W. H. Quantum computers, factoring, and decoherence. *Science* **270**, 1633–1635 (1995).
15. Freeman, R. *Spin Choreography* (Spektrum, Oxford, 1997).
16. Braunstein, S. L. *et al.* Separability of very noisy mixed states and implications for NMR quantum computing. *Phys. Rev. Lett.* **83**, 1054–1057 (1999).
17. Schack, R. & Caves, C. M. Classical model for bulk-ensemble NMR quantum computation. *Phys. Rev. A* **60**, 4354–4362 (1999).
18. Linden, N. & Popescu, S. Good dynamics versus bad kinematics: Is entanglement needed for quantum computation? *Phys. Rev. Lett.* **87**, 047901 (2001).
19. Coppersmith, D. *An Approximate Fourier Transform Useful in Quantum Factoring* (IBM Res. Rep. RC19642, IBM T. J. Watson Research Centre, Yorktown Heights, New York, 1994).
20. Vold, R. L. & Vold, R. R. Nuclear magnetic relaxation in coupled spin systems. *Prog. NMR Spectrosc.* **12**, 79–133 (1978).
21. Jeener, J. Superoperators in magnetic resonance. *Adv. Magn. Reson.* **10**, 1–51 (1982).
22. Bloch, F. Nuclear induction. *Phys. Rev.* **70**, 460–474 (1946).
23. Kraus, K. *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).
24. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
25. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
26. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
27. Vandersypen, L. M. K. *et al.* Implementation of a three-quantum-bit search algorithm. *Appl. Phys. Lett.* **76**, 646–648 (2000).
28. Aho, A. V., Sethi, R. & Ullman, J. D. *Compilers: Principles, Techniques and Tools* (Addison-Wesley, Reading, Massachusetts, 1986).
29. Green, M., Mayne, N. & Stone, F. G. A. Chemistry of the metal carbonyls. Part XLVI. Perfluorobutadienyl iron, rhenium and manganese complexes. *J. Chem. Soc. A* 902–905 (1968).
30. Leung, D. W. *et al.* Efficient implementation of selective recoupling in heteronuclear spin systems using Hadamard matrices. *Phys. Rev. A* **61**, 042310 (2000).

Acknowledgements

We thank X. Zhou and J. Preskill for discussions, J. Smolin for the use of his IBM workstation, D. Miller for help with spectral analysis, A. Schwartz and his team for their technical assistance, and J. Harris, W. Risk and H. Coufal for their support. L.V. acknowledges a Yansouni Family Stanford graduate fellowship. This work was supported in part by the QuARC project under a DARPA Quantum Information Science and Technology grant.

Correspondence and requests for materials should be addressed to I.L.C. (e-mail: ichuang@media.mit.edu).

A limit on spin–charge separation in high- T_c superconductors from the absence of a vortex-memory effect

D. A. Bonn[†], Janice C. Wynn[†], Brian W. Gardner[†], Yu-Ju Lin[†], Ruixing Liang^{*}, W. N. Hardy^{*}, J. R. Kirtley[‡] & K. A. Moler[†]

^{*} Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, V6T 1Z1 Canada

[†] Departments of Applied Physics and Physics, Stanford University, Stanford, California 94305, USA

[‡] IBM T.J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA

There is a long-standing debate about whether spin–charge separation is the root cause of the peculiar normal-state properties and high superconducting transition temperatures of the high- T_c materials. In the proposed¹ state of matter, the elementary excitations are not electron-like, as in conventional metals, but rather the electron ‘fractionalizes’ to give excitations that are chargeless spin-1/2 fermions (spinons) and charge $+e$ bosons (chargons). Although spin–charge separation has been well established in one dimension, the theoretical situation for two dimensions is controversial

and experimental evidence for it in the high- T_c materials is indirect. A model² with sharp experimental tests for a particular type of separation in two dimensions has recently been proposed. Here we report the results of those experimental tests, placing a conservative upper limit of 190 K on the energy of the proposed topological defects known as visons. There is still debate³ about the extent to which this experiment can settle the issue of spin–charge separation in the high- T_c copper oxides, because some forms of the separation are able to avoid the need for visons. But at least one class^{4–6} of theories that all predict a vortex-memory effect now are unlikely models for the copper oxides.

The model being tested is a version of spin–charge separation cast in the form of a Z_2 gauge theory with vortex-like topological defects called visons⁴. For the high-transition-temperature (high- T_c) superconductors, spin–charge separation could provide a route to superconductivity via the Bose condensation of chargons. In this respect, visons play an important role in the Z_2 gauge theory, ensuring that if Bose condensation of $+e$ chargons is indeed the route to superconductivity, the flux in magnetic vortices is still quantized in units of $\Phi_0 = h/2e$, as observed experimentally for the copper oxides^{7,8}. In conventional superconductors, where pairing is the route to superconductivity, a Cooper pair with charge $2e$ circling around an $h/2e$ vortex experiences a phase shift of 2π , but in the Z_2 gauge theory a chargon would pick up only a phase shift of π . In order for the chargon wavefunction to remain single-valued, it picks up an additional π phase shift from a vison that accompanies the vortex.

Senthil and Fisher have proposed an experimental test for visons in a copper-oxide superconductor fashioned into a ring geometry². The inset of Fig. 1 depicts such a ring, prepared with trapped

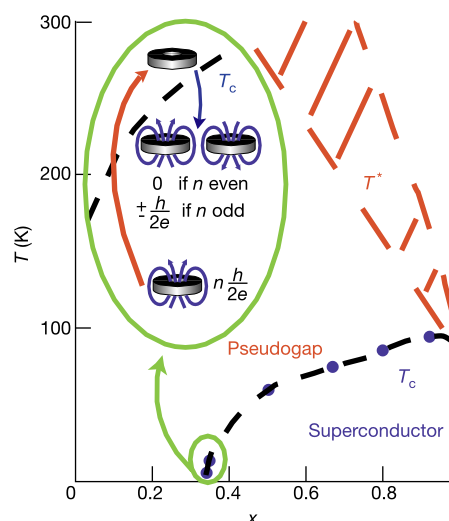


Figure 1 Vortex ‘memory’ in the $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ phase diagram. The superconducting transition temperature T_c can be tuned by adjusting the oxygen content x . The phase diagram also includes the characteristic temperature T^* , the temperature at which a crossover occurs to a region of peculiar behaviour associated with a pseudogap in the density of states. The lowest transition temperatures are found close to $x = 0.35$, produced by annealing the samples of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ near 900 °C in flowing oxygen. After this annealing, the samples are not superconducting, but annealing near room temperature for a few weeks under dry nitrogen allows the intercalated oxygen to order into Cu–O chains, producing samples with sharp superconducting transitions of 12 K or lower. The search for visons is concentrated in this low- T_c range. One predicted effect is the vortex ‘memory’ depicted in the inset. A superconducting ring containing one or more magnetic flux quanta is raised above T_c in zero magnetic field, but may still contain a vison if the starting point was an odd number of flux quanta. The presence of this vison would then spontaneously generate a vortex of random sign when the ring is recooled.

magnetic flux in the superconducting state. If the number of trapped flux quanta is odd, the ring will also contain a vison that could persist above T_c . If the ring is recooled in zero field before the vison escapes, it will spontaneously generate a single flux quantum of random sign. (A ring prepared with an even number of flux quanta would not contain a vison, and would not exhibit this 'vortex memory' after cycling the temperature.) We have seen no sign of the vortex-memory effect in the experiments reported here, placing a limit on the stability of visons and constraining the use of a class of spin-charge separation theories to explain the unusual properties of the copper oxides.

The experiment requires a ring with a T_c that is low enough to allow the sample temperature to be cycled faster than the vison escape time τ_v , expressed as

$$\tau_v = \tau_0 \exp[E_{\text{vison}}/(k_B T)] \quad (1)$$

where τ_0 is a microscopic attempt time, and E_{vison} is the energy

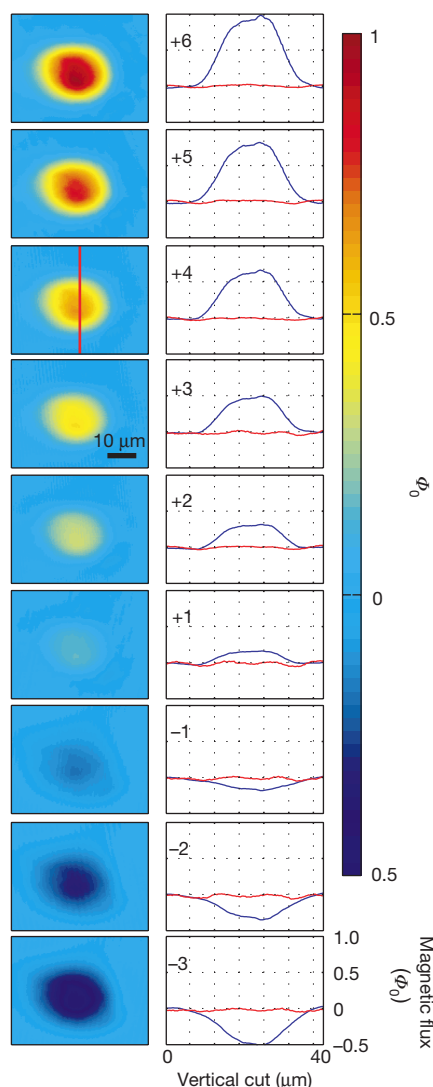


Figure 2 Magnetic flux trapped in a superconducting ring. The scanning SQUID magnetometer images to the left show different numbers ($+6\Phi_0$ to $-3\Phi_0$) of flux quanta trapped in a hole in a $\text{YBa}_2\text{Cu}_3\text{O}_{6.35}$ sample by cooling in a small magnetic field. The colour indicates the flux through a square ($8\mu\text{m} \times 8\mu\text{m}$) pickup loop. On the right, vertical cross-sections of the images show the magnetic flux in the hole before (blue) and after (red) cycling the ring's temperature in zero external magnetic field. In all cases, the trapped vortices escape in less than 10 s at 9.5 K.

barrier for vison escape⁴. Below T_c , the lifetime of a trapped flux quantum can be expressed similarly, with a characteristic energy E_{vortex} that includes a contribution from E_{vison} if the number of trapped quanta is odd. An underdoped sample with a low T_c means that the thermal energy $k_B T$ can be kept small throughout the thermal cycling. Low doping also helps by increasing E_{vison} , which is predicted⁴ to be the pseudogap energy scale $k_B T^*$ (Fig. 1).

The low- T_c samples of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ used in this study have improved crystalline quality, obtained by using BaZrO_3 crucibles for crystal growth⁹. Figure 1 shows the T_c of these high-purity crystals as a function of the oxygen content that sets the doping. The experiment was performed on a series of four crystals with successively lower transition temperatures, each having a single 10- μm -diameter hole fabricated with a focused ion beam. The trapped magnetic flux was imaged with a scanning SQUID magnetometer with a square ($8\mu\text{m} \times 8\mu\text{m}$) Nb pickup loop and a 70- μm scan range at 4 K (ref. 10). Three of the samples were cleaved into squares of roughly $50\mu\text{m} \times 50\mu\text{m}$, so that it was possible to image the entire sample and ensure that magnetic flux was not trapped outside the 10- μm hole. Both sample and magnetometer were mounted in a helium-flow cryostat surrounded by three layers of mu-metal shielding and an electromagnet used to apply small fields.

Figure 2 shows an ensemble of experiments to test for the vortex 'memory' associated with visons in the first of the small rings. The images on the left are of magnetic flux trapped in the 10- μm hole. The somewhat irregular shape is due to irregularities in the cleaved sample. The ring was prepared in each state by cooling in a small magnetic field, trapping between $+6$ and -3 flux quanta. After imaging, the sample was quickly heated and cooled again in zero external magnetic field. Subsequent imaging indicated that the flux escapes upon heating to 9.5 K for less than 10 s, and does not reappear. There was also no dependence on whether the starting condition involved an even or odd number of trapped flux quanta.

A more stringent search involved a pair of samples annealed to

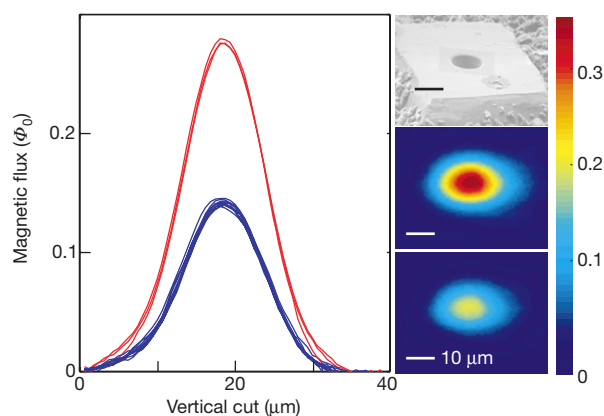


Figure 3 Cross-sections and images of flux quanta trapped in a ring with $T_c = 6.0$ K. The upper right image shows the typical geometry of the rings used in the experiment; they are squares ($50\mu\text{m} \times 50\mu\text{m}$) of single-crystal $\text{YBa}_2\text{Cu}_3\text{O}_{6.35}$, with a 10- μm hole drilled using a focused ion beam. The middle and lower right images show double ($2\Phi_0$) and single (Φ_0) flux quanta trapped in the ring. Integration of the total flux in the images confirms that the flux quantization is $\Phi_0 = h/2e$. The cross-sections shown on the left are obtained from 20-pixel-wide vertical stripes through the images, with the 20 pixels averaged in order to reduce noise and carefully confirm the reproducibility of the flux quanta. The many cross-sections show that single and double flux quanta (blue and red curves, respectively) could be repeatedly generated in the ring, although there were also numerous instances where some of the flux appeared to 'leak' into the annulus of the ring, a possible indication of granular behaviour. When the ring was heated in null field to only 5.6 K for 1 s, then recooled to 2 K, imaging indicated that these flux quanta escape with no sign of the vortex memory associated with visons.

have lower transition temperatures: one sample had $T_c \approx 7.5$ K, and the other had a T_c sufficiently low that a 6.0 K onset (2 K width) of the superconducting transition could be measured *in situ* with the scanning SQUID. A serious difficulty with the lowest- T_c sample was that magnetic flux was frequently trapped in the annulus of the ring rather than in the hole, and sometimes a single quantum of flux seemed to 'leak' from the hole into the annulus—a possible sign of granular behaviour, even though the rings are made from single crystals. The indications of granular behaviour only in this lowest- T_c ring probably stem from slight inhomogeneity in the oxygen content of the crystal, leading to inhomogeneous doping. If this is the case, then the material that constitutes the junctions between superconducting grains is slightly more underdoped than the material in the bulk of the grains. In the Senthil–Fisher picture, the intergrain material is a 'fractionalized insulator' and the vison gap would still exist in the barrier¹¹. So—despite this indication of granularity—when discrete flux quanta were clearly trapped in the hole, the temperature could be cycled in null field to test for memory effects. Figure 3 shows cross-sections and images of quantized flux in the ring with $T_c = 6.0$ K. Flux quanta of Φ_0 and $2\Phi_0$ were repeatedly trapped in the ring and then found to escape by heating to slightly below T_c , typically after only 1 s at 5.6 K, again with no sign of vortex memory.

The null results can be expressed as a limit on the energy of a vison by inserting the 1-s limit at 5.6 K into equation (1). The value of τ_0 is unknown, but a lower limit would be the time for an electron to cross one unit cell of the CuO_2 lattice. This estimate of $\tau_0 = 2 \times 10^{-15}$ s gives a conservative upper limit of $E_{\text{vison}} < 190$ K, below the predicted pseudogap energy scale, $k_B T^*$. A lower bound that is commonly used in discussions of vortex creep in conventional superconductors is $\tau_0 = 10^{-12}$ s, based on the characteristic frequencies of lattice vibrations¹². This higher estimate of τ_0 would give an upper limit of $E_{\text{vison}} < 160$ K.

Observations of static $h/2e$ fluxoids⁸ also set a low limit on E_{vison} in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, but are a less-strict test for visons because the $h/2e$ fluxoids could be metastable. In contrast, by measuring two steady states (before and after cycling the temperature), the Senthil–Fisher experiment that we report here can provide information on the dynamics, and hence the energy scale, of possible trapped visons². The upper limit of 190 K restricts any role that visons might have in the physics of these materials in the underdoped regime. Although the extent to which this experiment can settle the issue of spin–charge separation in the copper oxides is at present a matter of considerable debate³, at least one class of spin–charge separation theories^{4–6}—all of which predict a vortex memory effect—now seem to be unlikely models for the copper oxides. □

Received 9 August; accepted 19 October 2001.

- Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).
- Senthil, T. & Fisher, M. P. A. Fractionalization in the cuprates: Detecting the topological order. *Phys. Rev. Lett.* **86**, 292–295 (2001).
- Moessner, R., Sondhi, S. L. & Fradkin, E. Short-ranged RVB physics, quantum dimer models and Ising gauge theories. Preprint cond-mat/0103396 at (<http://xxx.lanl.gov>) (2001).
- Senthil, T. & Fisher, M. P. A. Z_2 gauge theory of electron fractionalization in strongly correlated electron systems. *Phys. Rev. B* **62**, 7850–7881 (2000).
- Franz, M. & Tesanovich, Z. Flux tubes, visons, and vortices in spin-charge separated superconductors. Preprint cond-mat/0010436 at (<http://xxx.lanl.gov>) (2000).
- Sedgewick, R. D., Scalapino, D. J. & Sugar, R. L. Fractionalized phase in an XY- Z_2 gauge model. Preprint cond-mat/0012028 at (<http://xxx.lanl.gov>) (2000).
- Gough, C. E. *et al.* Flux quantization in a high- T_c superconductor. *Nature* **326**, 855 (1987).
- Wynn, J. C. *et al.* Limits on spin-charge separation from $h/2e$ fluxoids in very underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. Lett.* **87**, 197002 (2001).
- Liang, R., Bonn, D. A., Hardy, W. N., Wynn, J. C. & Moler, K. A. Preparation and characterization of homogeneous YBCO single crystals with doping level near the SC-AFM boundary. *Physica C* (submitted).
- Gardner, B. W. *et al.* Scanning superconducting quantum interference device susceptometry. *Rev. Sci. Instrum.* **72**, 2361–2364 (2001).
- Senthil, T. & Fisher, M. P. A. Detecting fractions of electrons in the high- T_c cuprates. *Phys. Rev. B* **64**, 214511 (2001).
- Beasley, M. R., Labusch, R. & Webb, W. W. Flux creep in type II superconductors. *Phys. Rev.* **181**, 682–700 (1969).

Acknowledgements

We thank T. Senthil, M.P.A. Fisher, S. Sachdev, S. Kivelson, P.A. Lee, R.B. Laughlin and P.W. Anderson for discussions, and R. Lee for technical assistance. This work was supported by NSF CAREER and MRSEC grants, a Terman fellowship, the Natural Science and Engineering Research Council of Canada, the Canadian Institute for Advanced Research and the Institute for Theoretical Physics.

Correspondence and requests for materials should be addressed to D.A.B. (e-mail: bonn@physics.ubc.ca).

Folding-driven synthesis of oligomers

Keunchan Oh*, Kyu-Sung Jeong† & Jeffrey S. Moore*

* Roger Adams Laboratory, Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

The biological function of biomacromolecules such as DNA and enzymes depends on their ability to perform and control molecular association, catalysis, self-replication or other chemical processes. In the case of proteins in particular, the dependence of these functions on the three-dimensional protein conformation is long known¹ and has inspired the development of synthetic oligomers and polymers with the capacity to fold in a controlled manner^{2–7}, but it remains challenging to design these so-called 'foldamers' so that they are capable of inducing or controlling chemical processes and interactions^{8,9}. Here we show that the stability gained from folding can be used to control the synthesis of oligomers from short chain segments reversibly ligated through an imine metathesis reaction. That is, folding shifts the ligation equilibrium^{10–13} in favour of conformationally ordered sequences, so that oligomers having the most stable solution structures form preferentially. Crystallization has previously been used to shift an equilibrium in order to indirectly influence the synthesis of small molecules¹⁴, but the present approach to selectively prepare macromolecules with stable conformations directly connects folding and synthesis, emphasizing molecular function rather than structure in polymer synthesis.

Imine metathesis (Fig. 1a) was chosen as the ligation reaction because the reaction conserves bond energies and, under appropriate conditions, is a fast and reversible process at room temperature¹⁵. The equilibrium constant is close to unity; thus, there will be no inherent bias in the absence of an external driving force. Imine metathesis is slow in neutral organic solvents at room temperature¹⁶, which allows the reaction to be quenched and products to be analysed. In the presence of an appropriate acid catalyst, however, the reaction proceeds at a reasonable rate. Among ten different mono- and di-carboxylic acids screened, oxalic acid was found to be the most suitable as equilibrium was generally achieved in less than one hour at room temperature. In addition to reactivity considerations, we also noted that the geometry of the imine bond is commensurate with the compact helical conformation of the *meta*-connected phenylene ethynylene oligomers. This was confirmed by molecular modelling studies and solvent denaturation experiments as described below.

On the basis of Monte Carlo conformational searching, dodecamer **3e** (Fig. 1) was predicted to fold into a helical conformation with about six units per turn (see Supplementary Information). Experimentally, **3e** was observed to exhibit a conformational transition from a random to a helical state as the solvent changed from chloroform to acetonitrile. In particular, the ratio of ultraviolet absorbances at 303 and 289 nm (A_{303}/A_{289}) was used to monitor

† Present address: Department of Chemistry, Yonsei University, Seoul 120-749, Korea.

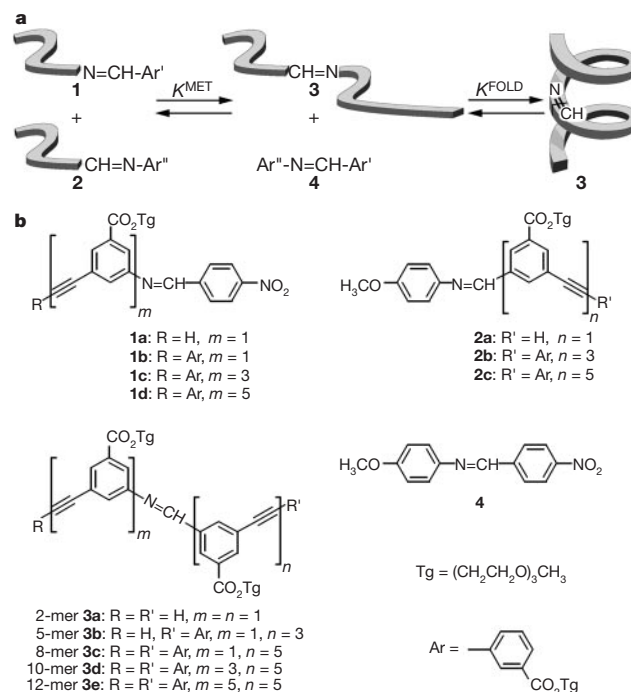


Figure 1 The imine metathesis of *meta*-connected phenylene ethynylene oligomers. **a**, Representation of the ligation reaction of oligo (*m*-phenylene ethynylene) imines. The coupled equilibria correspond to imine metathesis and helical folding. The folding equilibrium depends on the length of the ligated oligomer, the solvent, and the temperature. Equilibrium constants for the metathesis of aromatic imines are 1.0 ± 0.2

the cisoid population (that is, the fraction of repeat units in the helical conformation)^{5,17}. A plot of A_{303}/A_{289} versus solvent composition (expressed as a volume ratio) gave a sigmoidal curve with a midpoint at 63% CHCl₃, indicative of a cooperative conformational transition. The corresponding stabilizing energy ($-\Delta G$) of the helical conformation relative to the ensemble of random conformations was estimated¹⁸ to be 3.0 ± 0.2 kcal mol⁻¹. For an oligomer of the same size having no imine bonds, $-\Delta G$ is 3.2 ± 0.1 kcal mol⁻¹; thus, the imine unit destabilizes the helix very minimally, if at all.

To test the hypothesis that folding can drive a ligation reaction, two types of imine metathesis experiments were performed. One experiment was devised to determine how the metathesis equilibrium depends on the length of the oligomer chain synthesized under conditions that favour folding. If folding drives ligation, we expect that a critical size¹⁹ is necessary for folding-driven ligation to occur. The other experiment monitored the metathesis equilibrium under solvent conditions that favour folding and compared this equilibrium state to metathesis reactions occurring in solvents that do not favour folding.

Data in entries 1–5 of Table 1 were first collected in CD₃CN to examine the question of the chain length. For example, reaction

regardless of substituents¹⁶. **b**, Chemical structure and corresponding compound numbers of various imines used in this study. Oligomer segments of different lengths (**1a–d** and **2a–c**) were synthesized by repeated palladium/copper-catalysed coupling of appropriate aryl halide and terminal acetylene precursors^{17,20}. *N*-terminal and *C*-terminal imine groups were added by condensation of the corresponding aldehydes and amines.

between *N*-terminal dimer **1b** and *C*-terminal hexamer **2c** in CD₃CN at 294 K with oxalic acid as catalyst produces the ligation product **3c** (entry 3, Table 1). Because **3c** is too short to fold we did not expect to see significant equilibrium shifting. As shown in Fig. 2a, new resonance signals corresponding to the imine CH protons of exchange products **3c** and **4** appear in the ¹H NMR spectra, and their intensities increase asymptotically as equilibrium is attained. The molar ratio of reactants to products remains constant after about an hour (Fig. 2c), and from the integrated intensities, the equilibrium constant (K_{eq}) was estimated to be approximately 2.6 for this reaction.

In contrast, the metathesis between a pair of longer oligomers under otherwise identical conditions exhibited considerable equilibrium shifting. For example, *N*- and *C*-terminal hexamer imines, **1d** and **2c** undergo metathesis to dodecamer **3e**, an oligomer capable of folding into a helix of roughly two turns (entry 5 in CD₃CN, Table 1; Fig. 2b). Plots of the reaction quotient, *Q*, versus time again reveal that an asymptotic limit is reached from which an equilibrium constant of about 62 was determined. Examination of entries 1–5 in CD₃CN in Table 1 supports the notion of a critical chain length. For example, equilibrium constants are approximately

Table 1 Equilibrium constants for various imine metathesis reactions

Entry	Reactants		Ligation product	Approx. extent of helicity*	K_{eq} (294 K)†		$\Delta\Delta G$ (kcal mol ⁻¹)‡
					CD ₃ CN	CDCl ₃	
1	1a	2a	3a (2-mer)	None	1.1	1.2 ± 0.1	0.00
2	1a	2b	3b (5-mer)	None	1.1 ± 0.1	1.4 ± 0.2	0.00
3	1b	2c	3c (8-mer)	1 1/3 turns	2.6 ± 0.1	1.6 ± 0.1	-0.50
4	1c	2c	3d (10-mer)	1 2/3 turns	19 ± 1	1.9 ± 0.1	-1.66
5	1d	2c	3e (12-mer)	2 turns	62 ± 2	1.9 ± 0.1	-2.35

* Assumes 6 units per turn.

† Measured by ¹H NMR. Error estimates are based on the standard deviation of values measured in duplicate runs.

‡ $\Delta\Delta G$ is a measure of equilibrium shifting expressed as the difference in free energy for formation of entry *x* relative to dimer formation (entry 1) in CD₃CN. $\Delta\Delta G_x = \Delta G(\text{entry } x) - \Delta G(\text{entry } 1) = -RT \ln K_{eq} + RT \ln 1.1$.

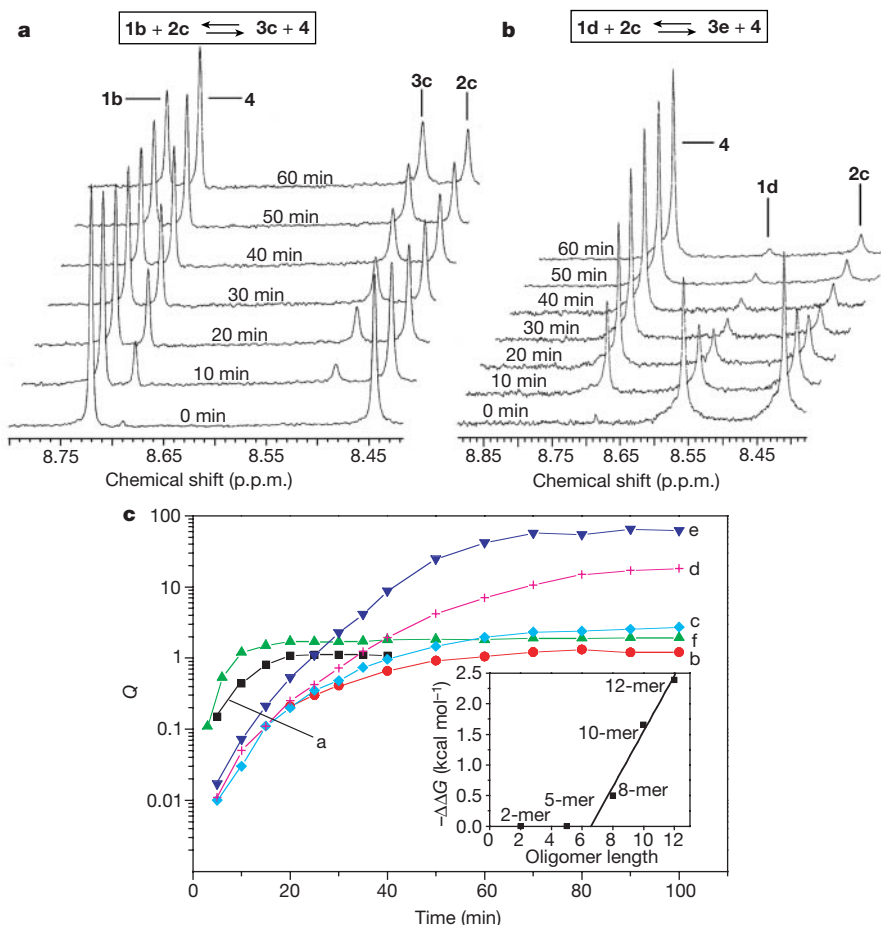


Figure 2 The partial ^1H NMR (400 MHz) spectra showing the imine $\text{CH}=\text{N}$ region for the substrate and products of the imine metathesis reaction. **a**, Signals from the reaction of N -terminal dimer **1b** and C -terminal hexamer **2c** as a function of the reaction time. The reaction was performed in CD_3CN with 5 mol% of oxalic acid relative to total imine at 21°C , and an initial concentration of 5 mM for both **1b** and **2c**. A small extent ($<10\%$) of imine hydrolysis was observed during metathesis due to the presence of adventitious water in solution. Because the metathesis proceeds much faster than the hydrolysis under the conditions, the latter does not interfere in determining the equilibrium. **b**, Signals from the reaction of N -terminal hexamer **1d** and C -terminal hexamer **2c** as a function of the reaction time. The reaction was performed in CD_3CN with 5 mol% of oxalic acid relative to total imine at 21°C , and an initial concentration of 5 mM for both **1d** and **2c**. Under the reaction conditions, the NMR spectrum of ligation product **3e** is broad and poorly resolved

due to product aggregation and the signal for its imine proton is not observed. **c**, Plots of Q versus time for imine metatheses; trace a, **1a** + **2a** in CD_3CN , trace b, **1a** + **2b** in CD_3CN , trace c, **1b** + **2c** in CD_3CN , trace d, **1c** + **2c** in CD_3CN , trace e, **1d** + **2c** in CD_3CN , and trace f, **1d** + **2c** in CDCl_3 . Here, Q is the reaction quotient given by $\prod_i C_i^{\nu_i}$ where C_i is the concentration of species i and the ν_i represent stoichiometric coefficients. The molar ratios were determined from peak heights of imine ^1H NMR signals (assignments made by comparison to authentic imines, prepared separately by condensations of the corresponding aldehydes and amines). Inset, the magnitude of equilibrium shifting ($\Delta\Delta G$, kcal mol^{-1}) as measured by the free-energy difference between ΔG (entry x in Table 1) and ΔG (entry 1 in Table 1) in CD_3CN as a function of the chain length of the ligated oligomers.

2.6 and 19 for metatheses of C -terminal hexamer **2c** with N -terminal dimer **1b** and tetramer **1c**, respectively (Fig. 2c and Table 1). When magnitudes of equilibrium shifting, expressed in terms of $\Delta\Delta G$ relative to dimer formation, are plotted against the length of ligation products, evidence for a critical chain length is apparent. The short chains have no shift in equilibrium state while there is a linear free-energy relationship for the octamer through the dodecamer (Fig. 2c inset). The x -intercept occurs just above the hexamer, suggesting that the critical chain length is about one turn. The linear correlation between $\Delta\Delta G$ and chain length above this critical value is consistent with the notion that the folding stability is directly proportional to the contact area between adjacent turns, as expected for a compact helical conformation.

The reactants listed in entries 1–5 of Table 1 were then equilibrated in CDCl_3 to study the process under solvent conditions that do not favour folding. In particular, when the metathesis reaction between **1d** and **2c** is conducted in CDCl_3 instead of CD_3CN , an equilibrium constant of about 1.9 was obtained (Table 1, entry 5). This behaviour is consistent with the observation that the dodeca-

mer **3e** folds into a helical conformation in acetonitrile, but it exists in a random conformational state in chloroform. The large equilibrium shifting observed in acetonitrile for longer oligomers can thus be attributed to the additional stabilizing energy gained from the folding process. We do not completely exclude aggregation as contributing to the driving force for equilibrium shifting but the evidence shows that an intramolecular folding process is dominant. Light-scattering data on **3e** in acetonitrile shows only monodisperse particles ranging from 1.0 nm to 2.0 nm, which is roughly the size of a single chain in the helical conformation. The fact that only oligomers longer than a critical chain length have considerable equilibrium shifting suggests that equilibrium shifting depends on folding. Moreover, in a set of independent experiments we have found that the product ratio changes by less than 10% as the concentration of reactants changes by two orders of magnitude (from 50 to 0.5 mM). The possibility that the metathesis product **4** might template the reaction was excluded because the imine chemical shifts of **4** did not vary for entries 1–5 of Table 1 and integrated intensities accounted for complete mass balance of these

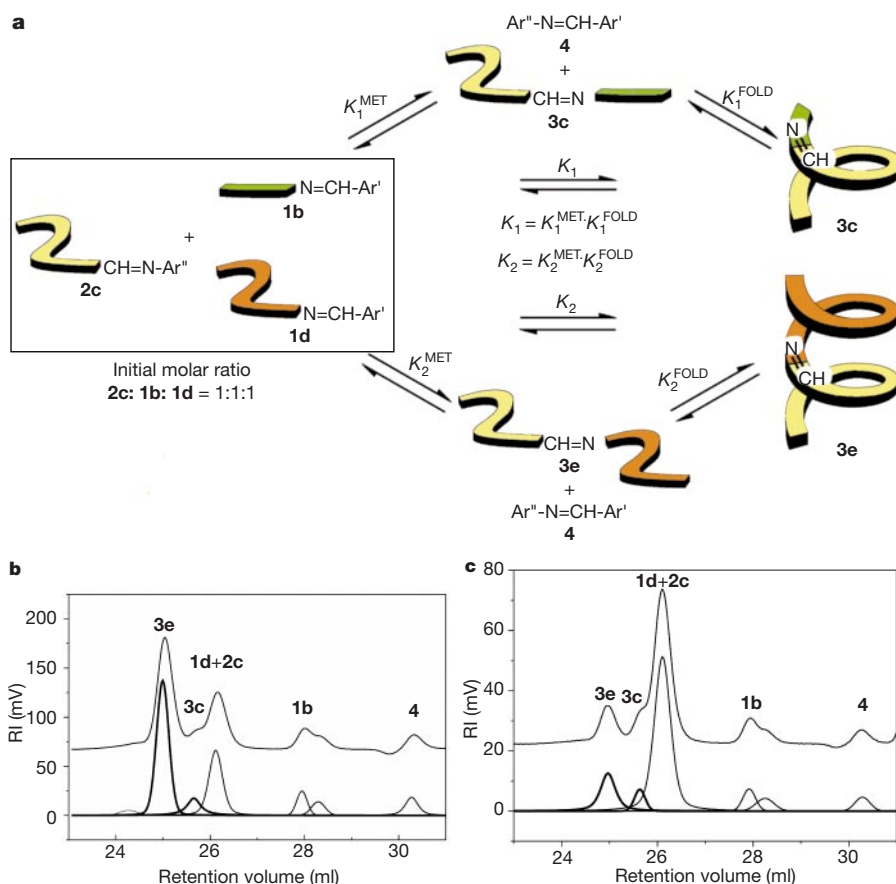


Figure 3 Segment selection in imine metathesis determined by foldability. **a**, Representation of the competitive ligation of oligo(*m*-phenylene ethynylene) imines. K_1^{MET} is the equilibrium constant for the imine metathesis between *N*-terminal dimer **1b** and *C*-terminal hexamer **2c**. K_2^{MET} is the equilibrium constant for the imine metathesis between *N*-terminal hexamer **1d** and *C*-terminal hexamer **2c**. K_1^{FOLD} and K_2^{FOLD} are equilibrium constants for the respective helix-coil transitions resulting in the overall equilibrium constants K_1 and K_2 . **b**, Gel permeation chromatogram of the reaction mixture from imine metathesis between equimolar quantities of *N*-terminal dimer **1b**, *N*-terminal hexamer **1d**, and *C*-terminal hexamer **2c**. The reaction was performed in CH_3CN with 3 mol% of oxalic acid relative to total imine at 21 °C, and an initial concentration of 5 mM

in each of **1b**, **1d** and **2c**. The upper trace is the experimentally observed chromatogram and the lower trace is the deconvolution. **c**, Same as **b** except in $CHCl_3$. A small extent of imine hydrolysis was observed during metathesis due to presence of adventitious water in solution. Gel permeation chromatography measurements were performed in tetrahydrofuran at 21 °C with a Waters 515 (high-performance liquid chromatography) pump, Viscotek TDA model 300 triple detector, and a series of three Viscotek 7.8 × 300 mm columns (2 × GMHXL16141 and 1 × G3000HXL16136) which were calibrated with narrow molecular mass polystyrene standards over the range 1,100 to 4.5×10^6 dalton with a column exclusion limit of 1.0×10^7 dalton. Flow rate was 1.0 ml min⁻¹. RI is a refractive index.

equilibria. We confirmed that the reverse imine metathesis between **3e** and **4** occurred in CD_3CN . The overall equilibrium constants are consistent with $K = K^{MET}K^{FOLD}$ where K^{MET} , the equilibrium constant for imine metathesis, is about 1 for all reactions and K^{FOLD} , the equilibrium constant for helix-coil transitions, is obtained from the solvent denaturation experiments.

The next hypothesis tested was that folding could be used to select, from a pool of oligomers, those segments most prone to fold. For competitive ligation (Fig. 3a), the product distribution will depend on the relative stability of all components. As the longer oligomer can gain a greater stabilizing energy from folding, this ligation product is predicted to be preferentially selected over the shorter oligomer in a competitive ligation reaction. From the two equilibrium expressions of the individual ligation reactions and the independently determined equilibrium constants $K_1 = 2.6$ and $K_2 = 62$ in CD_3CN (from Table 1), we calculate the $[3e]/[3a]$ ratio to be 5.6. (K_1 and K_2 are defined in Fig. 3a.)

To test this prediction, two *N*-terminal imines (**1b** and **1d**) in equimolar quantities were allowed to compete for one equivalent of *C*-terminal imine **2c**. The product of **1b** + **2c** is the octamer **3c**, while **1d** and **2c** produce the dodecamer **3e**. These segment selection experiments were conducted under both folding (CH_3CN) and non-folding ($CHCl_3$) conditions. A difference in the distribution of

metathesis products reflects the preferential synthesis of those oligomers capable of folding. The product distribution was probed by both gel permeation chromatography and ¹H NMR. Gel permeation chromatograms of the equilibrium mixtures show a marked shift in the distribution of ligation products under folding versus non-folding conditions. In CH_3CN , where the longer ligation product is in the compact helical state, the molar ratio of dodecamer **3e** to octamer **3c** is about 3.6 (Fig. 3b). In $CHCl_3$, where both ligation products are in their random conformational states, this ratio is 1.6 (Fig. 3c). Similar results were obtained when the reaction was monitored by ¹H NMR spectroscopy. In CD_3CN , the amount of reactant **1d** consumed was three times the amount of **1b** to make the longer ligation product **3e**. However, in $CDCl_3$, ¹H NMR showed that **1b** and **1d** were equally consumed, in agreement with the independently measured equilibrium constants of $K_1 = 1.6$, $K_2 = 1.9$ (see Supplementary Information). Our findings are now being extended to the size-selective synthesis of folded oligomers by dynamic templating. □

Received 21 July; accepted 8 November 2001.

1. Mirsky, A. E. & Pauling, L. On the structure of native, denatured, and coagulated proteins. *Proc. Natl. Acad. Sci. USA* **22**, 439–447 (1936).
2. Gellman, S. H. Foldamers: A manifesto. *Acc. Chem. Res.* **31**, 173–180 (1998).
3. Nowick, J. S. Chemical models of protein β -sheets. *Acc. Chem. Res.* **32**, 287–296 (1999).

4. Katz, T. J. Syntheses of functionalized and aggregating helical conjugated molecules. *Angew. Chem. Int. Edn Engl.* **39**, 1921–1923 (2000).
5. Nelson, J. C., Saven, J. G., Moore, J. S. & Wolynes, P. G. Solvophobically driven folding of nonbiological oligomers. *Science* **277**, 1793–1796 (1997).
6. Hirschberg, J. H. K. K. *et al.* Helical self-assembled polymers from cooperative stacking of hydrogen-bonded pairs. *Nature* **407**, 167–170 (2000).
7. Berl, V., Huc, I., Khoury, R. G., Kirsche, M. J. & Lehn, J.-M. Interconversion of single and double helices formed from synthetic molecular strands. *Nature* **407**, 720–723 (2000).
8. Kirshenbaum, K., Zuckermann, R. N. & Dill, K. A. Designing polymers that mimic biomolecules. *Curr. Opin. Struct. Biol.* **9**, 530–535 (1999).
9. Stigers, K. D., Soth, M. J. & Nowick, J. S. Designed molecules that fold to mimic protein secondary structures. *Curr. Opin. Chem. Biol.* **3**, 714–723 (1999).
10. Lehn, J.-M. & Eliseev, A. V. Dynamic combinatorial chemistry. *Science* **291**, 2331–2332 (2001).
11. Lehn, J.-M. Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chem. Eur. J.* **5**, 2455–2463 (1999).
12. Huc, I. & Lehn, J.-M. Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl Acad. Sci. USA* **94**, 2106–2110 (1997).
13. Moore, J. S. & Zimmerman, N. W. “Masterpiece” copolymer sequences by targeted equilibrium-shifting. *Org. Lett.* **2**, 915–918 (2000).
14. Woodward, R. B., Sondheimer, F., Taub, D., Heusler, K. & McLarmore, W. M. The total synthesis of steroids. *J. Am. Chem. Soc.* **74**, 4223–4251 (1952).
15. Trost, B. M. The atom economy—A search for synthetic efficiency. *Science* **254**, 1471–1477 (1991).
16. Töth, G., Pinter, I. & Messmer, A. Mechanism of the exchange reaction of aromatic Schiff bases. *Tetrahedr. Lett.* **15**, 735–738 (1974).
17. Prince, R. B., Saven, J. G., Wolynes, P. G. & Moore, J. S. Cooperative conformational transitions in phenylene ethynylene oligomers: Chain-length dependence. *J. Am. Chem. Soc.* **121**, 3114–3121 (1999).
18. Pace, C. N., Shirley, B. A. & Thomson, J. A. in *Protein Structure: A Practical Approach* (ed. Creighton, T. E.) 311–330 (IRL, New York, 1989).
19. Goodman, M., Verdini, A. S., Toniolo, C., Phillips, W. D. & Bovey, F. A. Sensitive criteria for the critical size for helix formation in oligopeptides. *Proc. Natl Acad. Sci. USA* **64**, 444–450 (1969).
20. Zhang, J., Pesak, D. J., Ludwick, J. L. & Moore, J. S. Geometrically-controlled and site-specifically-functionalized phenylacetylene macrocycles. *J. Am. Chem. Soc.* **116**, 4227–4239 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank C. F. Zukoski and A. Y. Mirarefi for discussions and comments on light-scattering experiments. This work was supported by the National Science Foundation and the US Department of Energy, Division of Materials Sciences, through the Frederick Seitz Materials Research Laboratory at the University of Illinois at Urbana-Champaign.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.M. (e-mail: moore@scs.uiuc.edu).

Striped iron zoning of olivine induced by dislocation creep in deformed peridotites

J. Ando*, Y. Shibata*, Y. Okajima*, K. Kanagawa†, M. Furusho‡ & N. Tomioka§

* Department of Earth and Planetary Systems Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan

† Department of Earth Sciences, Chiba University, Chiba 263-8522, Japan

‡ OYO Corporation, 10-20 Hamasaki Avenue, Hyogo-ku, Kobe 652-0807, Japan

§ Department of Earth and Planetary Sciences, Kobe University, Kobe 657-8501, Japan

Deformation of solid materials affects not only their microstructures, but also their microchemistries^{1–5}. Although chemical unmixing of initially homogeneous multicomponent solids is known to occur during deformation by diffusion creep^{4,5}, there has been no report on their chemical zoning due to deformation by dislocation creep, in either natural samples or laboratory experiments. Here we report striped iron zoning of olivine

((Mg,Fe)₂SiO₄) in deformed peridotites, where the iron concentration increases at subgrain boundaries composed of edge dislocations. We infer that this zoning is probably formed by alignment of edge dislocations dragging a so-called Cottrell ‘atmosphere’ of solute atoms^{3,6,7} (iron in this case) into subgrain boundaries during deformation of the olivine by dislocation creep. We have found that the iron zoning does not develop in laboratory experiments of high strain rates where dislocations move too fast to drag the Cottrell atmosphere. This phenomenon might have important implications for the generation of deep-focus earthquakes, as transformation of olivine to high-pressure phases preferentially occurs in high-iron regions, and therefore along subgrain boundaries which would be preferentially aligned in plastically deformed mantle peridotites.

We have done microstructural and microchemical analyses of olivine grains in deformed mantle-derived peridotites from the Uenzaru peridotite complex in the Hidaka metamorphic belt of central Hokkaido, Japan. Olivine grains develop a lattice preferred orientation (LPO) characterized by a [100] density maximum subparallel to lineation, as commonly observed in deformed peridotites⁸. Most olivine grains in our samples exhibit undulose extinction or kink bands due respectively to gradual or sharp intragranular misorientation (Fig. 1). It is known that such microstructures are formed by alignment of dislocations into planar arrays to form subgrain boundaries due to crystal plastic deformation. Oxidation decoration⁹ and subsequent observation by scanning electron microscopy¹⁰ revealed that zones of undulose extinction in fact correspond to subgrain boundaries. In addition, transmission electron microscopy (TEM) revealed that subgrain boundaries are composed of edge dislocations with (010)[100] or {0kl}[100] slip, and that dislocation densities are high—of the order of 10⁷–10⁸ cm^{–2} (data not shown). These dislocation microstructures, as well as developed LPO, indicate that the olivine grains in our samples have been deformed by dislocation creep.

Striped zoning of Mg and in particular of Fe is easily visible in olivine grains that exhibit undulose extinction: the zoning is revealed by mapping of element concentrations using electron-probe microanalysis. The regions of low Mg and high Fe concentrations are exactly parallel to zones of undulose extinction—that is, subgrain boundaries (Fig. 2). Measurements of Fe concentration across subgrain boundaries by analytical transmission electron microscopy (ATEM) clearly indicate that Fe concentration increases at subgrain boundaries (Fig. 3). Areas of high Fe concentration correspond to those of low Mg concentration (Table 1), indicating that the striped zoning is formed via interdiffusion between Mg and

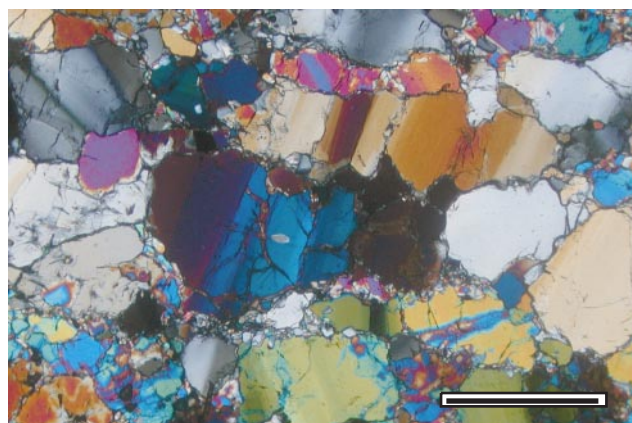


Figure 1 Optical micrograph in crossed polarized light, showing undulose extinction and kink bands (NNE–SSW direction) in olivine grains of an Uenzaru peridotite. Such olivine microstructures are commonly developed in deformed peridotites derived from the Earth's upper mantle. Scale bar, 1 mm.

Fe. Both electron-probe microanalysis and ATEM measurements indicate that the concentrations of fayalite (Fe_2SiO_4) components at subgrain boundaries are 1–1.5% higher than those in the other regions (Fig. 3 and Table 1). Such striped Fe zoning of olivine as described here is also found in deformed peridotites from the Horoman peridotite complex of the Hidaka metamorphic belt, as well as in those from the Baldissero, Balumuccia and Finero peridotite complexes in the Ivrea zone of the southern Alps. The striped Fe zoning of olivine may therefore be common in mantle peridotites.

Dislocation cores that give rise to elastic distortion of the crystal lattice attract solute atoms in alloys and minerals, forming Cottrell atmospheres^{3,6,7}. Olivine is a solid solution between Mg_2SiO_4 (forsterite) and Fe_2SiO_4 (fayalite), in which Fe^{2+} is a solute cation. ATEM measurements³ of X-ray intensities of Fe across a free edge dislocation in olivine from a mantle peridotite have revealed that Fe concentration increases in the immediate vicinity (less than $0.5\ \mu\text{m}$) of the dislocation, indicating a Cottrell atmosphere of solute Fe. During dislocation creep of a crystal, dislocation glide and climb enable dislocations to align in planar arrays and form subgrain boundaries such that the internal strain energy of the crystal is minimized. The formation of subgrain boundaries composed of edge dislocations dragging a Cottrell atmosphere of Fe can therefore produce a striped Fe zoning of olivine.

Two possible post-deformational processes may also affect the striped Fe zoning via pipe diffusion of Fe along dislocations: subsolidus reaction^{11,12} and metasomatism¹³. The sources of Fe in these two cases are clinopyroxene and an Fe-rich fluid phase, respectively. However, the striped Fe zoning of olivine is developed even in dunite and harzburgite in which clinopyroxene is absent. In addition, Fe-enrichment at the rims of olivine grains possibly in direct contact with fluid was not detected. Laihunite, an Fe^{3+} -bearing olivine, has been reported to be formed as 0.6-nm-thick layer parallel to (001) along dislocation cores and grain boundaries of olivine during metasomatism of mantle peridotites under upper-mantle conditions^{13,14}. But laihunite and any other precipitates were not found along subgrain boundaries by our TEM observations. An additional synchrotron micro-XANES study (not shown) did not detect Fe^{3+} in the iron-rich or in the iron-poor areas of the olivine grains that we studied (XANES; X-ray absorption near edge structure). The possibility of mantle metasomatism is therefore excluded, and consequently, post-deformational processes cannot be responsible for the observed striped Fe zoning of olivine.

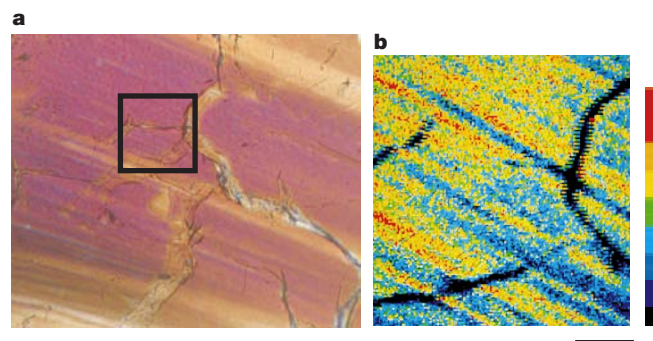


Figure 2 Microstructure and microchemistry of a strongly deformed olivine grain, indicating the relationship between undulose extinction and Fe zoning. **a**, Optical micrograph in crossed polarized light, showing undulose extinction in the WNW–ESE direction. **b**, Map of Fe concentrations within the rectangle shown in **a**. Striped Fe zoning in the WNW–ESE direction can be seen. Reddish colours indicate higher Fe concentrations (see colour bar on right of panel). Scale bar (black, at bottom), $30\ \mu\text{m}$. Black lines cutting the striped Fe zoning in **b** are cracks. Note that high-Fe areas are exactly parallel to zones of undulose extinction, that is, subgrain boundaries.

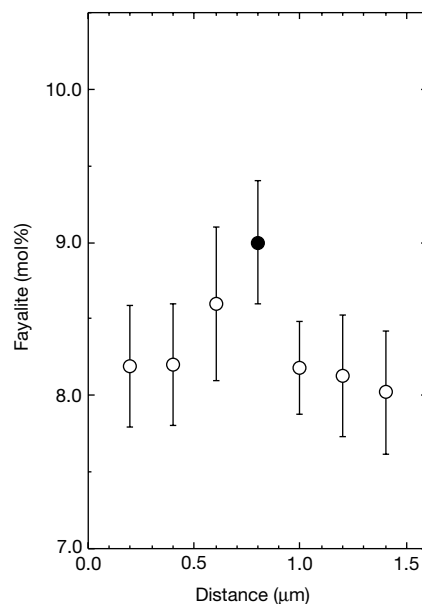


Figure 3 Fayalite (Fe_2SiO_4) components measured by ATEM across a subgrain boundary composed of edge dislocations. The filled circle shows a measurement made just at the subgrain boundary. The Fe concentration at the subgrain boundary is higher than that in the adjacent areas.

The (010)[100] and $\{0kl\}[100]$ slip systems in olivine observed in the Uenzaru peridotites are active at relatively high temperatures—for example, at temperatures higher than 900°C when the confining pressure and strain rate are $1,500\ \text{MPa}$ and $10^{-4}\ \text{s}^{-1}$, respectively¹⁵. Furusho and Kanagawa¹⁶ have suggested that the Uenzaru peridotites are deformed at temperatures of 760 – 960°C and pressures of 700 – $1,000\ \text{MPa}$. These ranges of temperatures and pressures correspond to those in the uppermost mantle and in descending slabs^{17–19}.

Subgrain boundaries composed of edge dislocations are perpendicular to the slip direction, that is, the Burgers vector. TEM observations (not shown) of dislocations in the studied olivine grains indicate the slip direction to be [100], and therefore subgrain boundaries are parallel to (100). In fact, optically visible subgrain boundaries in olivine grains are mostly parallel to their (100) planes.

Table 1 Chemical composition of olivine

	Fe-poor region			Fe-rich region		
SiO_2	40.91	40.90	41.09	40.77	40.48	40.94
TiO_2	0.02	—	—	0.01	0.02	0.02
FeO	8.35	8.00	7.66	9.22	9.14	8.92
NiO	0.33	0.40	0.37	0.34	0.36	0.36
Cr_2O_3	0.00	0.04	0.03	0.01	0.03	0.02
MnO	0.10	0.14	0.16	0.18	0.15	0.14
MgO	50.71	49.52	50.25	49.70	49.92	49.61
CaO	0.02	—	0.02	—	—	—
Total	100.44	99.00	99.58	100.23	100.10	100.01
No. of cations (O=4)						
Si	0.993	1.005	1.002	0.996	0.990	1.000
Ti	0.000	—	—	0.000	0.000	0.000
Fe	0.169	0.164	0.156	0.188	0.187	0.182
Ni	0.006	0.008	0.007	0.007	0.007	0.007
Cr	0.000	0.001	0.001	0.000	0.001	0.000
Mn	0.002	0.003	0.003	0.004	0.003	0.003
Mg	1.835	1.814	1.827	1.809	1.821	1.806
Ca	0.001	—	0.001	—	—	—
Total	3.006	2.995	2.997	3.004	3.009	2.998
Fa*	8.4	8.3	7.9	9.4	9.3	9.2

Compositions are given in wt%. Dash indicates under detection limit ($<0.01\ \text{wt}\%$). All measurements are within a single olivine grain. Fa* = $100 \times \text{Fe}/(\text{Mg} + \text{Fe})$.

LPO with a [100] density maximum close to lineation indicates preferred alignment of (100) and hence subgrain boundaries normal to lineation. This has an important implication for the generation of deep-focus earthquakes, which are possibly triggered by transformation of olivine to high-pressure phases^{19–24}. In a descending slab, the phase transformation of olivine to β -phase (wadsleyite) and γ -phase (ringwoodite) would occur first at regions of high Fe concentrations, because Fe-rich olivine transforms at a lower pressure than Mg-rich olivine at the same temperatures^{25,26}. Moreover, stacking faults along which the crystal lattice is transformed to ringwoodite are preferentially formed parallel to (100) in olivine^{27,28}. Thus the phase transformation would preferentially occur parallel to (100) in olivine, which is preferentially aligned normal to lineation, that is, in planar arrays. The high-pressure phases thus formed are much finer-grained than the host olivine^{21,23,24}, and possibly deform superplastically by a grain-size-sensitive flow—in contrast to the host olivine plastically deforming by grain-size-insensitive dislocation creep. The formation of high-pressure phases in planar arrays may therefore lead to shear instability in descending slabs, and trigger deep-focus earthquakes in these slabs.

Orowan's equation indicates that strain rate is proportional to dislocation velocity²⁹. The observed Fe zoning in naturally deformed olivine implies the moving of dislocations at speeds slow enough to drag a Cottrell atmosphere, probably because of very low strain rates ($\sim 10^{-14} \text{ s}^{-1}$) in the mantle. If the strain rate is much higher, as in deformation experiments (generally $\geq 10^{-6} \text{ s}^{-1}$), fast-moving dislocations would leave the Cottrell atmosphere behind, resulting in no Fe concentrations along dislocations. The presence of a Cottrell atmosphere also has a pinning effect on dislocations and inhibits their movement. Accordingly, flow laws at high and low strain-rate conditions are different³⁰. Hence the plasticity of olivine at very low strain rates in the mantle may be different from that at the high strain rates achieved in the laboratory—this needs to be explored in the future, to obtain a better understanding of the dynamics of the Earth's interior. □

Received 23 August; accepted 6 November 2001.

- Li, J. C. M., Nolfi, F. V. Jr & Johnson, C. A. Diffusional equilibrium of substitutional atoms in a stressed solid. *Acta Metall.* **19**, 749–752 (1971).
- Green, H. W. II On the thermodynamics of non-hydrostatically stressed solids. *Phil. Mag.* **41**, 637–647 (1980).
- Kitamura, M., Matsuda, H. & Morimoto, N. Direct observation of the Cottrell atmosphere in olivine. *Proc. Jpn Acad.* **62**, 149–152 (1986).
- Dimos, D., Wolfenstine, J. & Kohlstedt, D. L. Kinetic demixing and decomposition of multi-component oxides due to a nonhydrostatic stress. *Acta Metall.* **36**, 1543–1552 (1988).
- Ozawa, K. Stress-induced Al–Cr zoning of spinel in deformed peridotites. *Nature* **338**, 141–144 (1989).
- Cottrell, A. H. in *Report of the Strength of Solids* 30–36 (The Physical Society, London, 1948).
- Blavette, D., Cadel, E., Fraczkiewicz, A. & Menand, A. Three-dimensional atomic-scale imaging of impurity segregation to line defects. *Science* **286**, 2317–2319 (1999).
- Ben Ismail, W. & Mainprice, D. An olivine fabric database: An overview of upper mantle fabrics and seismic anisotropy. *Tectonophysics* **296**, 145–157 (1998).
- Kohlstedt, D. L., Goetze, C., Durham, W. B. & Vandersande, J. B. New technique for decorating dislocations in olivine. *Science* **19**, 1045–1091 (1976).
- Karato, S. Scanning electron microscope observation of dislocations in olivine. *Phys. Chem. Minerals* **14**, 245–248 (1986).
- Ozawa, K. Partitioning of elements between constituents minerals in peridotites from the Miyamori ultramafic complex, Kitakami Mountains, Northeast Japan: Estimation of P–T condition and igneous composition of minerals. *J. Fac. Sci. Univ. Tokyo II* **21**, 115–137 (1986).
- Kawasaki, T. & Ito, E. An experimental determination of the exchange of Fe^{2+} and Mg^{2+} between olivine and Ca-rich clinopyroxene. *Am. Mineral.* **79**, 461–477 (1994).
- McGuire, A. V., Dyar, M. D. & Nielson, J. E. Metasomatic oxidation of upper mantle peridotite. *Contrib. Mineral. Petrol.* **109**, 252–264 (1991).
- Banfield, J. F., Dyar, M. D. & McGuire, A. V. The defect microstructure of oxidized mantle olivine from Dish Hill, California. *Am. Mineral.* **77**, 977–986 (1992).
- Carter, N. L. & Ave Lallemand, H. G. High temperature flow of dunite and peridotite. *Geol. Soc. Am. Bull.* **81**, 2181–2202 (1970).
- Furusho, M. & Kanagawa, K. Transformation-induced strain localization in a lherzolite mylonite from the Hidaka metamorphic belt of central Hokkaido, Japan. *Tectonophysics* **313**, 411–432 (1999).
- Ito, E. & Sato, H. in *High-Pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghni, H.) 257–262 (Terrapub/American Geophysical Union, Tokyo/Washington DC, 1992).
- Iidaka, T. & Furukawa, Y. Double seismic zone for deep earthquakes in the Izu-Bonin subduction zone. *Science* **263**, 1116–1118 (1994).
- Kirby, S. H., Stein, S., Okal, E. A. & Rubie, D. C. Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Rev. Geophys.* **34**, 261–306 (1996).

- Kirby, S. H. Localized polymorphic phase transformations in high-pressure faults and applications to the physical mechanism of deep earthquakes. *J. Geophys. Res.* **92**, 13789–13800 (1987).
- Green, H. W. II & Burnley, P. C. A new self-organizing mechanism for deep-focus earthquakes. *Nature* **341**, 733–737 (1989).
- Kirby, S., Durham, W. B. & Stern, L. A. Mantle phase changes and deep-earthquake faulting in subducting lithosphere. *Science* **252**, 216–225 (1991).
- Burnley, P. C., Green, H. W. II & Prior, D. J. Faulting associated with the olivine to spinel transformation in Mg_2GeO_4 and its implications for deep focus earthquakes. *J. Geophys. Res.* **96**, 425–443 (1991).
- Tingle, T. N., Green, H. W. II, Scholtz, C. H. & Kocynski, T. A. The rheology of faults triggered by the olivine–spinel transformation in Mg_2GeO_4 and its implications for the mechanism of deep focus earthquakes. *J. Struct. Geol.* **15**, 1249–1256 (1993).
- Katsura, T. & Ito, E. The system $\text{Mg}_2\text{SiO}_4 - \text{Fe}_2\text{SiO}_4$ at high pressure and temperatures: precise determination of stabilities of olivine, modified spinel, and spinel. *J. Geophys. Res.* **94**, 15663–15670 (1989).
- Akaogi, M., Ito, E. & Navrotsky, A. Olivine-modified spinel-spinel transitions in the system $\text{Mg}_2\text{SiO}_4 - \text{Fe}_2\text{SiO}_4$: calorimetric measurements, thermochemical calculation, and geophysical. *J. Geophys. Res.* **94**, 15671–15685 (1989).
- Poirier, J.-P. in *Anelasticity in the Earth* (eds Stacey, F. D., Paterson, M. S. & Nicholas, A.) 113–117 (American Geophysical Union and Geological Society of America, Washington DC/Boulder, 1981).
- Kerschhofer, L., Rubie, D. C., Sharp, T. G., McConnell, J. D. C. & Dupas-Bruzek, C. Kinetics of intracrystalline olivine–ringwoodite transformation. *Phys. Earth Planet. Inter.* **121**, 59–76 (2000).
- Poirier, J.-P. in *Creep of Crystals* 62–63 (Cambridge Univ. Press, Cambridge, 1985).
- Yoshinaga, H. in *Rheology of Solids and of the Earth* (eds Karato, S. & Toriumi, M.) 29–41 (Oxford Science Publications, New York, 1989).

Acknowledgements

We thank K. Ozawa for critical reading of the manuscript, H. Ishisako for making thin sections, Y. Takahashi for his contribution to the XANES study at the KEK photon factory in Tsukuba, Japan, and N. Abe for supplying the specimens from the southern Alps. This study was supported by JSPS (J.A.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A. (e-mail: ando@letitbe.geol.sci.hiroshima-u.ac.jp).

Horses damp the spring in their step

Alan M. Wilson*, M. Polly McGuigan*, Anne Su† & Anton J. van den Bogert†

* Department of Veterinary Basic Sciences, The Royal Veterinary College, Hatfield, Herts AL9 7TA, UK

† Department of Biomedical Engineering, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA

The muscular work of galloping in horses is halved by storing and returning elastic strain energy in spring-like muscle–tendon units^{1,2}. These make the legs act like a child's pogo stick that is tuned to stretch and recoil at 2.5 strides per second. This mechanism is optimized by unique musculoskeletal adaptations: the digital flexor muscles have extremely short fibres and significant passive properties, whereas the tendons are very long and span several joints^{3,4}. Length change occurs by a stretching of the spring-like digital flexor tendons rather than through energetically expensive length changes in the muscle⁵. Despite being apparently redundant for such a mechanism⁵, the muscle fibres in the digital flexors are well developed. Here we show that the mechanical arrangement of the elastic leg permits it to vibrate at a higher frequency of 30–40 Hz that could cause fatigue damage to tendon and bone. Furthermore, we show that the digital flexor muscles have minimal ability to contribute to or regulate significantly the 2.5-Hz cycle of movement, but are ideally arranged to damp these high-frequency oscillations in the limb.

In a 500-kg horse, about 1,000 J of elastic energy are stored in the digital flexor tendons and suspensory ligament (interosseus muscle) of each leg in each stride^{2,6}. This is achieved by gravitational and

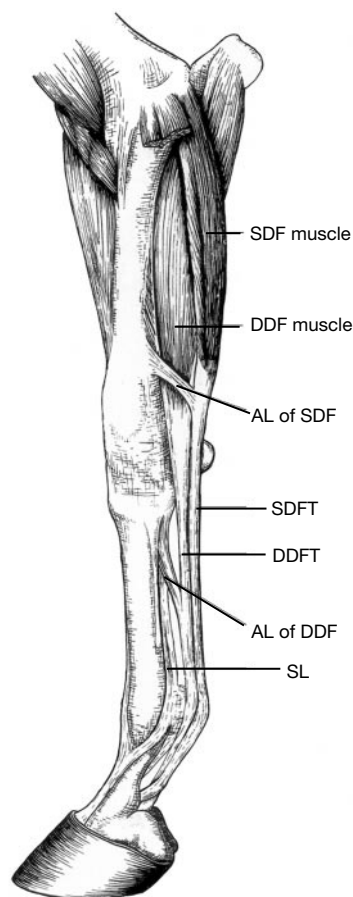


Figure 1 Equine distal forelimb (medial view) showing segments below the elbow, and the tendons and muscles that resist compression of the limb by the effect of gravitational and inertial forces during the stance phase of locomotion. This part of the limb is about 1 m long in a 450-kg thoroughbred racehorse. AL, accessory ligament; DDF, deep digital flexor; SDF, superficial digital flexor.

inertial forces that extend the metacarpophalangeal joint and stretch these collagenous structures (Fig. 1). Animals (including humans) that store elastic energy in their tendons have compliant tendons⁵ and minimize length change in the associated muscle^{7–9}. The muscle instead preloads^{10,11} and/or powers the system by shortening at low tendon forces⁷. Further efficiency gains may be achieved by shortening or removing the muscle fibres, because the energetic cost of generating isometric force is a function of muscle fibre length over activated volume^{3–5}. The endpoint of this optimization is shown in the completely collagenous interosseus muscle of the adult horse.

However, the equine forelimb superficial and deep digital flexor muscles remain well developed (mass 600–1,000 g), and 98% (see Methods) of the physiological cross-sectional area comprises extremely short (superficial, 2–6 mm; deep, 6–17 mm) muscle fibres^{1,6,10,11}. Similar short-fibred muscles exist in the hindlimb. Each muscle has a maximum isometric force of about 5,000 N ($n = 5$). The tendons are about 700 mm long and the muscles about 400 mm long. A peak tendon strain of 8–12% during locomotion^{1,12,13} equates to a tendon elongation of 70 mm and, assuming similar aponeurosis strains, a muscle elongation of 40 mm. Both muscles are protected from overextension by short (50 mm) and hence, when loaded, stiff accessory ligaments that link the tendon distal to the muscle belly to the bone, effectively acting as an additional parallel elastic element.

Assuming a change in sarcomere length of 35% and muscle force in excess of the long-fibred head isometric capacity (230 N), then

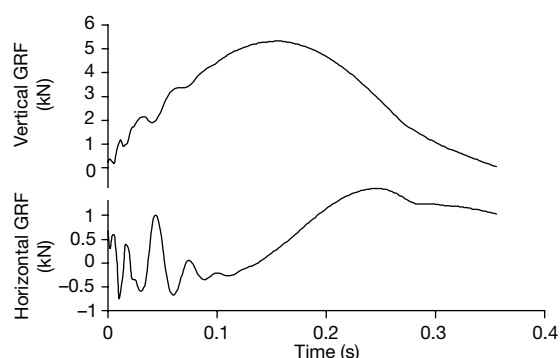


Figure 2 Plot of horizontal (in direction of locomotion) and vertical limb ground reaction force (GRF) against time for a 450-kg horse trotting at 3 m s^{-1} over a 6-mm-thick polyester/rubber-matting-covered conveyor-belt force plate. The inflections in the curves during early stance are due to limb vibration excited by foot impact.

the superficial and deep digital flexor muscle contractile elements should change length by only 1–2 mm and 2–6 mm, respectively. The long-fibred head appears sufficient to flex the digit during limb protraction, but the role of the short-fibred heads remains unclear. With potential contractile element length changes of 2 and 6 mm in a muscle tendon unit elongation of 60–80 mm, the muscles appear inappropriate for tuning the leg spring¹⁰ or driving the spring system as in other energy-storing tendons^{7–9}. It is possible, however, that the short fibres can, owing to lateral force transmission through the complex three-dimensional (3D) architecture of the muscle, achieve a length change greater than that predicted from fibre-length measurements¹⁴.

Hoof impact, slide and resonance lasts 20–30 ms. Foot impact excites the equine limb to vibrate in a cranio-caudal direction at a frequency of 30–40 Hz (Fig. 2)¹⁵. The vibrations are similar on different surfaces (tarmac $35.9 \pm 1.2 \text{ Hz}$; concrete $35.8 \pm 2.1 \text{ Hz}$; rubber matting $34.9 \pm 1.2 \text{ Hz}$; $n = 8$), and correspond to a muscle–tendon unit length change of about 2 mm (from joint angle changes) and an energy of 4–8 J. The vibrations are damped within 100 ms on hard surfaces and more quickly on soft surfaces. Similar frequency oscillations occur in the hindlimb and a simulation thereof⁶, but are damped within a cycle. Notably, 40–50-Hz horizontal vibrations are present for about 100 ms after impact in accelerometer and kinematic data recorded from the human tibia during barefoot running¹⁷.

Excessive high-frequency vibration in the limb will cause fatigue damage in both bone¹⁸ and tendon¹⁹ by increasing the number of loading cycles and the loading rate experienced by the distal limb tissues. Musculoskeletal injuries are extremely common in athletic horses, and most of them occur in this spring system, which has been shown to fail through cyclical fatigue damage in as little as 10,000 strides of gallop^{20,21}. A statistical correlation between high-frequency components in the ground reaction force and the development of tendonitis in racehorses has also been shown²².

Control and damping of vibration is an important design feature of manmade mechanical systems. Machines are usually designed to be rigid, but elastic deformation is an essential functional property of the equine limb and some innovative legged robots²³. This elasticity and the multi-joint leg spring makes limb vibration inevitable, and it seems essential that this vibration is appropriately damped. The digital flexor muscles have extremely short fibres but a large physiological cross-sectional area. Muscles are good at absorbing energy²⁴, so these short-fibred muscles appear ideally arranged for absorbing energy during high-frequency small displacement vibration.

One front leg from each of six horses killed at an abattoir was mounted in a hydraulic press at three different limb orientations representing impact through to mid-stance. The limbs vibrated for

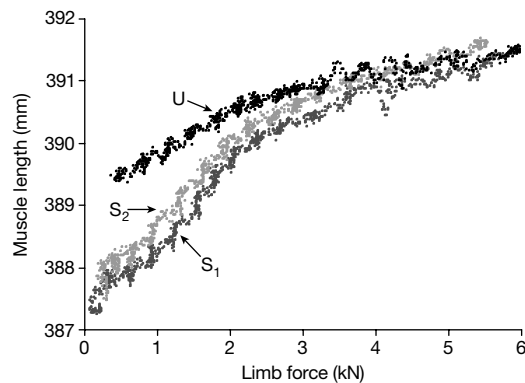


Figure 3 Relationship between limb force and muscle length for an equine superficial digital flexor muscle loaded *in situ* in the unstimulated state (U), and two sequential cycles with maximal electrical stimulation (S_1 , S_2). The difference between plots U and S_1 , S_2 represents the length change in the contracted muscle, and the difference between plots S_1 and S_2 represents the decline in contractile properties with repeated loading. Similar data were obtained for the deep digital flexor muscle.

3–5 cycles at 25–45 Hz when an impulse was applied to the carpal joint with a 1-kg hammer. Vibration frequency increased with limb force (28 Hz at 1 kN to 38 Hz at 4 kN) and was 4 Hz higher at impact orientation (26.5° to vertical) than at mid-stance orientation. Another six legs were compressed up to 6 kN (roughly gallop force) at mid-stance orientation over 5 s, twice with and twice without electrical stimulation (50 Hz, 60 V) of the digital flexor muscles. It is unlikely that full muscle activation is achieved in this loading system (or during locomotion), but the length change would still be similar at low limb forces. Experiments were completed within 1 h of death, muscle temperature remained above 37°C, and repeat plots were similar (Fig. 3).

At low limb forces, both muscles were 2–3 mm shorter when stimulated. The stimulated muscle lengthened more than the unstimulated muscle with increasing limb force (Fig. 3), absorbing energy in the limb force range at which the leg vibrations are damped (Fig. 2). The stimulated muscle lengthened more because the muscle was shorter at the start of loading, and when the isometric capacity of the contractile element was exceeded the muscle lengthened and the load was transferred to the parallel elastic element and the accessory ligament.

The accessory ligament is shorter (50 mm compared with 300 mm) and hence stiffer (N mm^{-1}) than the muscle aponeurosis, so at high elongations most of the force will pass through it. This load transfer from the compliant muscle to the stiff accessory ligament accounts for the plateau of both the active and the passive muscle length curves above a limb force of 3 kN. The accessory ligament therefore limits muscle length change, keeping the muscle fibres at an appropriate length for developing active force and hence absorbing energy during oscillation.

Muscle stimulation had minimal effect on the relationship between limb force and limb length, showing that there is minimal change in limb stiffness and thus little potential for limb tuning through contraction of the digital flexor muscles. The limb spring may, however, be tuned by the muscles associated with the elbow and shoulder joints or those attaching the scapula to the trunk.

We created a numerical model of the system that reproduced published *in vivo* joint angles and tendon strains (see Methods). The model in the reference state (leg vertical, no muscle activation and tarmac ground surface) also reproduced similar low (2.9 Hz) and high (57.5 Hz) natural frequencies to *in vivo* and *ex vivo* data when loaded with a mass of 200 kg, that is, half body mass (Fig. 4). Reducing vertical and horizontal surface stiffness by 500% from that representing tarmac ($k = 5 \times 10^5 \text{ N m}^{-1}$) to that representing

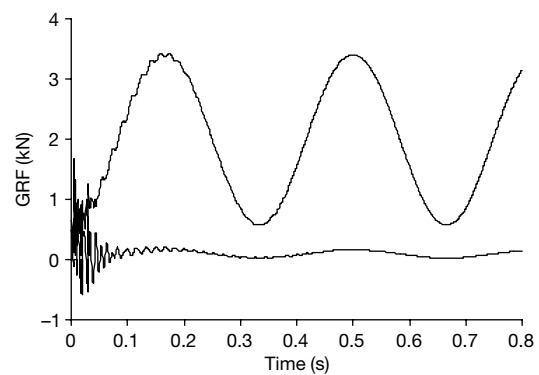


Figure 4 Vertical and horizontal ground reaction force (GRF) data produced from the equine limb model described in the text in the reference condition (leg vertical, no muscle activation, loading on tarmac). Top trace, vertical data; bottom trace, horizontal data. The initial oscillations are generated by the energy of the model being loaded by a mass of 200 kg. Two frequencies of oscillation are apparent at 2.9 and 57.5 Hz. There is no muscle stimulation in this case, and hence vibration damping is limited to the properties of the foot surface.

grass ($k = 1 \times 10^5 \text{ N m}^{-1}$) reduced the resonance frequencies by a much smaller amount to 2.7 and 34.1 Hz. This is due to the reduction in stiffness of the overall system. An increase in damping coefficient increased damping of vibration but did not change the frequency. The model was loaded at limb orientations of foot impact (26.5° to vertical) and mid-stance (vertical). At impact orientation, the resonance frequencies dropped from reference to 2.2 and 53.2 Hz because the limb is more compliant at this orientation.

In the model, muscle activation created muscle length changes greater than observed in the limb-loading experiment, had a damping effect on low-frequency oscillation (limb energy storage), but only a 3% tuning effect (2.9 to 3 Hz), which is less than the surface effect. There was, however, no damping or tuning effect on high-frequency limb vibration. This is because Hill muscle models (almost universal in biomechanical modelling) fail to simulate high-frequency energy absorption²⁴, because at high frequencies (and hence velocities) the contractile element remains almost isometric and the series elastic element absorbs most length change. Real muscles from a diverse range of animals (crayfish, frog, rat and rabbit), however, show a plateau or minimum in stiffness around 30–60 Hz (refs 24, 25). This effect is absent from muscle in rigor²⁵. This suggests that the minimum is an inherent property of cross-bridge dynamics²⁵, and we would predict that horse muscle will have similar frequency-dependent properties.

The cross-bridge model^{26,27} gives excellent simulations of the forces generated by live muscle fibres in response to small rapid length changes. We used the model to predict the energy absorbed by a cross-bridge when subjected to sinusoidal length changes of 0.5% at a range of frequencies. The assumption of constant displacement is reasonable because of the stiff accessory ligament in series with the muscle. There was a broad peak energy absorption of $15 \times 10^{-21} \text{ J}$ per cross-bridge per cycle centred around 100 Hz (Fig. 5), which is similar to that reported for real muscle^{24,25}. At 40 Hz, a single cross-bridge would absorb $12 \times 10^{-21} \text{ J}$ per cycle. Scaling-up for the digital flexor muscles (1 kg of muscle containing 0.3 mM myosin heads) gives a value of 2.5 J per cycle, which is sufficient to provide useful damping of limb vibration (energy 4–8 J), but is tiny compared with the energetic cost of locomotion². At stride frequency, the energy absorption was only $6 \times 10^{-21} \text{ J}$ per cross-bridge.

Muscle fibres will therefore absorb more vibrational energy around 30–100 Hz than predicted by the Hill model, in which stiffness increases monotonically with frequency. Experimentally

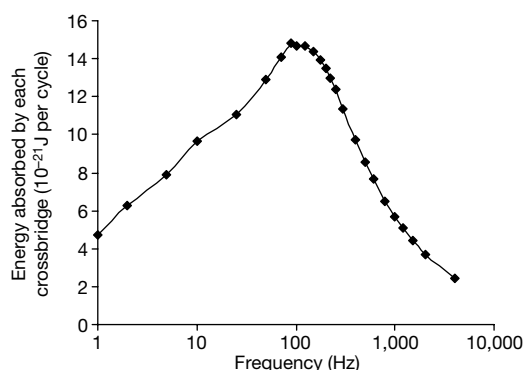


Figure 5 Plot of cycle work done on each cross-bridge in 10^{-21} J per cycle as a function of frequency (Hz) during a sinusoidal length change of 0.5%. These data were generated using the cross-bridge model of refs 26 and 27.

derived frequency-dependent muscle stiffness and phase data²⁵ were used to predict the impulse response of a muscle–tendon–mass system geared to represent the limb vibration mechanism. We found resonance at 30 Hz and 88% damping in the first 100 ms of the impulse response. This simple model shows a similar response to that predicted from the cross-bridge simulations and supports the hypothesis that the muscles have appropriate architecture and properties to damp vibrations in the frequency range that occur in the equine limb.

In summary, the apparently functionless but well-developed digital flexor muscles in the horse appear to act as dampers of high-frequency limb vibration rather than to flex the digit or tune the leg spring. A cross-bridge muscle model can reproduce this muscle energy absorption and should be useful in recreating this high-frequency damping in musculoskeletal models of impact attenuation. □

Methods

In vivo measurements

Eight horses were trotted over a forceplate (Kistler 9287BA) topped with either 6-mm-thick conveyor belt matting (a composite of polyester fibre and artificial rubber), bituminous roadstone (tarmac) or concrete. The frequency of leg vibration was determined from each plot of ground reaction force/time and averaged for six runs (three per forelimb) per horse and surface.

Measurements of muscle architecture

The superficial digital flexor muscle and the humeral (long-, intermediate- and short-fibred components separate), ulnar and radial heads of the deep digital flexor muscle were dissected free. We weighed each head and measured the fascial length (along the line of fibres between aponeuroses) at ten sites distributed across each muscle head. We calculated physiological cross-sectional area assuming a muscle density of 1.1 g cm^{-3} , and maximum isometric force assuming a capacity of 30 N cm^{-2} . No further correction for pennation angle or collagenous volume of the muscle was made.

Ex vivo limb loading

Forelimbs were cut perpendicular to the limb axis just proximal to the olecranon with the elbow at a mid-stance angle and a 13-mm diameter hole drilled through the distal humerus and the elbow joint. The leg was mounted in a hydraulic jig through a flat plate with a central, hinged 12-mm pin placed into the pre-drilled hole. The foot was placed on a force-measuring foot plate. We placed retroreflective markers over the centres of rotation of the elbow, carpal, metacarpophalangeal and distal interphalangeal joints. Retroreflective headed pins were inserted into the origin and muscle tendon junction of the deep and superficial digital flexor muscles, and data were recorded at 240 Hz (ProReflex, Qualisys). Wire electrodes (128-strand) were threaded through the deep and superficial digital flexor muscles, which were then wrapped in plastic film to prevent evaporative cooling.

Musculoskeletal model

A two-dimensional skeletal model of the equine forelimb was constructed, using five rigid-body segments connected by four ideal hinges. We used published segmental parameters (length, mass, moment of inertia and locations of the centre of mass)²⁸, and our model was conceptually similar to that in ref. 16. SD/FAST software (PTC) was used to obtain

equations of motion. The model was stabilized by two muscle–tendon units, the superficial and deep digital flexors, two passive tendinous structures, the suspensory ligament and accessory ligament of the deep digital flexor, and a linear torsion spring, the palmar carpal ligament. Muscles were represented by a three-component Hill model based on that in ref. 29. Passive elastic elements were represented by a quadratic force–length relationship. Ground contact forces F as a function of deformation x were applied at two points on the hoof using a visco-elastic model³⁰:

$$F = -k(x + b|\dot{x}|)$$

where $k = 5 \times 10^5$ or $1 \times 10^5 \text{ N m}^{-1}$, and $b = 0.5 \text{ s m}^{-1}$. Ground contact was modelled as a rigid hoof penetrating into a deformable surface. x is the penetration depth. The model coefficients were chosen such that this model accounts for the combined deformation of surface and hoof. The stiffness coefficient k for tarmac ($5 \times 10^5 \text{ N m}^{-1}$) is based on a peak ground reaction force of 5,000 N producing a static deformation of 1 cm; the equivalent deformation for turf ($k = 1 \times 10^5 \text{ N m}^{-1}$) is 5 cm. The damping coefficient b is based on ref. 30; $b = 0.5$ will produce a 50% increase in impact force when impact occurs at a speed of 1 m s^{-1} , when compared with static deformation.

Frequency domain model

The limb is represented by a one-dimensional muscle–tendon–mass system, representing the mechanism of limb vibration. With external force as the input and mass displacement as the output of the system, the transfer function $H(f)$ of this system in the frequency domain is:

$$H(f) = G \frac{k + s_f}{ks_f - 4(k + s_f)\pi^2 mf^2} \quad G = 3.0, k = 10^6 \text{ N m}^{-1}, m = 10 \text{ kg}$$

where f is the excitation frequency, k is tendon stiffness, and m is the oscillating mass, that is, the limb between the elbow and foot²⁸. G is the gear ratio for the limb, that is, the ratio between mass displacement and change in musculotendon length, derived from moment arms and bone lengths. The frequency-dependent stiffness s_f of the muscle, represented as a complex number with magnitude and phase, was taken from ref. 25 and scaled to a maximal force of 5,000 N and a fibre length of 5 mm. Four such muscle–tendon units were assumed to account for individual heads and accessory ligaments. Complex stiffness is the ratio between force and displacement when both are sinusoidal signals $e^{i\omega t}$ in the complex domain, where t is time and $i = \sqrt{-1}$. Stiffness is a complex number $S e^{i\phi}$ where ϕ is the phase difference between force and displacement and S is the amplitude ratio between force and displacement. The impulse response of the system was obtained by inverse Fourier transform of $H(f)$.

Received 1 March; accepted 12 October 2001.

- Dimery, N. J., Alexander, R. McN. & Ker, R. F. Elastic extension of leg tendons in the locomotion of horses (*Equus caballus*). *J. Zool.* **210**, 415–425 (1986).
- Minetti, A. E., Ardigo, L. P., Reinach, E. & Saibene, F. The relationship between mechanical work and energy expenditure of locomotion in horses. *J. Exp. Biol.* **202**, 2329–2338 (1999).
- Alexander, R. McN. & Bennet-Clark, H. C. Storage of elastic strain energy in muscle and other tissues. *Nature* **265**, 114–117 (1977).
- Alexander, R. McN. *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, Cambridge, 1988).
- Biewener, A. A. & Roberts, T. J. Muscle and tendon contributions to force, work and elastic energy savings: a comparative perspective. *Exercise Sport Sci. Rev.* **28**, 99–107 (2000).
- Biewener, A. A. Muscle–tendon stresses and elastic energy storage during locomotion in the horse. *Comp. Biochem. Physiol. B* **120**, 73–87 (1998).
- Biewener, A. A., Konieczynski, D. D. & Baudinette, R. V. *In vivo* muscle force–length behaviour during steady state hopping in tammar wallabies. *J. Exp. Biol.* **201**, 1681–1694 (1998).
- Hof, A. L. *In vivo* measurement of the series elasticity release curve of human triceps surae muscle. *J. Biomech.* **31**, 793–800 (1998).
- Kurokawa, S., Fukunaga, T. & Fukashiro, S. Behaviour of fascicles and tendinous structures of human gastrocnemius during vertical jumping. *J. Appl. Physiol.* **90**, 1349–1358 (2001).
- Grande, J. Penniform muscles of the horses forelimb. *J. Anat. Abstr.* **132**, 318 (1981).
- Hermanson, J. W. & Cobb, M. A. Four forearm flexor muscles of the horse, *Equus caballus*: anatomy and histochemistry. *J. Morphol.* **212**, 269–280 (1992).
- Stephens, P. R., Nunamaker, D. M. & Butterweck, D. M. Application of a Hall effect transducer for the measurement of tendon strains in horses. *Am. J. Vet. Res.* **50**, 1089–1095 (1989).
- Herrick, W. C., Kingsbury, H. B. & Lou, D. Y. S. A study of the normal range of strain, strain rate and stiffness of tendon. *J. Biomed. Mat. Res.* **12**, 877–884 (1978).
- Bogert, A. J., van den Gertsen, K. G. M. & Cole, G. K. Human muscle modeling from a user's perspective. *J. Electromyogr. Kinesiol.* **8**, 119–124 (1998).
- van Weeren, P. R., van den Bogert, A. J., Back, W., Bruin, G. & Barneveld, A. Kinematics of the standardbred trotter measured at 6, 7, 8 and 9 m s^{-1} on a treadmill after 5 months of prerace training. *Acta Anat.* **146**, 154–161 (1993).
- van den Bogert, A. J. & Schamhardt, H. C. Multi-body modelling and simulation of animal locomotion. *Acta Anat.* **146**, 95–102 (1993).
- Lake, M. J., Coyle, V. R. & Lees, A. High frequency characteristics of the lower limb during running. *Proc. 18th Congr. Int. Soc. Biomech.* (eds Müller, R., Gerber, H. & Stacoff, A.) 200–201 (Laboratory for Biomechanics, ETH Zurich, 2001).
- Carter, D. R. Mechanical loading histories and cortical bone remodeling. *Calcif. Tissue Int.* **36**(Suppl.), 19–24 (1984).
- Wang, X. T., Ker, R. F. & Alexander, R. McN. Fatigue rupture of wallaby tail tendons. *J. Exp. Biol.* **198**, 847–852 (1995).
- Cheney, J. A., Liou, S. Y. & Wheat, J. D. Cannon bone fracture in the thoroughbred racehorse. *Med. Bio. Eng.* **11**, 613–620 (1973).

21. Nunamaker, D. M., Butterweck, D. M. & Provost, M. T. Fatigue fractures in thoroughbred racehorses: relationships with age, peak bone strain, and training. *J. Orthop. Res.* **8**, 604–611 (1990).
22. Dow, S. M., Leendertz, J. A., Silver, I. A. & Goodship, A. E. Identification of subclinical tendon injury from ground reaction force analysis. *Equine Vet. J.* **23**, 266–272 (1991).
23. Pratt, G. A. *et al.* Stiffness isn't everything. *Proc. 4th Int. Symp. Experimental Robotics, ISER, Stanford, California* 173–178 (1995).
24. Ettema, G. J. C. & Huijing, P. A. Frequency response of rat gastrocnemius medialis in small amplitude vibrations. *J. Biomech.* **27**, 1015–1022 (1994).
25. Kawai, M. & Brandt, P. W. Sinusoidal analysis: A high resolution method for correlating biochemical reactions with physiological processes in the activated skeletal muscles of rabbit, frog and crayfish. *J. Muscle Res. Cell. Motil.* **1**, 279–303 (1980).
26. Piazzesi, G. & Lombardi, V. A cross-bridge model that is able to explain mechanical and energetic properties of shortening muscle. *Biophys. J.* **68**, 1966–1979 (1995).
27. Piazzesi, G. & Lombardi, V. Simulation of the rapid regeneration of the actin–myosin working stroke with a tight coupling model of muscle contraction. *J. Muscle Res. Cell. Motil.* **17**, 45–53 (1996).
28. van den Bogert, A. J. *Computer Simulation of Locomotion in the Horse*. Thesis, Univ. Utrecht, The Netherlands (1989).
29. van Soest, A. J. & Bobbert, M. F. The contribution of muscle properties in the control of explosive movements. *Biol. Cybern.* **69**, 195–204 (1993).
30. Gerritsen, K. G. M., van den Bogert, A. J. & Nigg, B. M. Direct dynamics simulation of the impact phase in heel–toe running. *J. Biomech.* **28**, 661–668 (1995).

Acknowledgements

We acknowledge the assistance of R. C. Woledge in carrying out the cross-bridge model calculations and thank him for comments on the manuscript. We thank the Horserace Betting Levy Board, London, for funding this work.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.W. (e-mail: awilson@rvc.ac.uk).

Maternal control of resting-egg production in *Daphnia*

Victor Alekseev* & Winfried Lampert

Max Planck Institute of Limnology, Department of Physiological Ecology,
Postfach 165, D-24302 Ploen, Germany

Many planktonic organisms produce 'resting' stages when the environmental conditions deteriorate. Like seeds, resting stages can survive unfavourable conditions. The crustacean *Daphnia* normally reproduces by means of parthenogenetically produced normal, not resting, eggs—but occasionally switches to bisexual reproduction, which results in two resting eggs encased in a robust structure carried on the back of the female. This 'ephippium' is shed with the next moult, and can survive dormant for many years. The induction of resting-egg production requires multiple environmental stimuli, one of them being photoperiod^{1,2}. The switch from production of parthenogenetic eggs to resting eggs in *Daphnia* has recently been shown to be influenced by a maternal food effect³. Here we present evidence that female *Daphnia* transmit information not only about food but also on photoperiod to their offspring, and influence the production of resting eggs in the next generation. The combined maternal effects can be relevant for the correct timing of resting-egg production—for example, in discriminating between spring and autumn conditions.

After the work on calanoid copepod resting eggs that can survive for 300 years (ref. 4), studies on egg banks (analogous to seed banks)

focused on *Daphnia*^{5,6}. Understanding the dynamics of the egg bank requires information on the production and hatching of resting eggs. Ehippia production in *Daphnia* displays a seasonal pattern. The induction of resting eggs has been long studied¹, but there is still no generally accepted theory about the causation of the seasonal pattern.

The switch in reproductive mode and the production of resting eggs require time, so it would be advantageous if the mother could anticipate the environmental conditions that her offspring will face. Growing interest in the control of phenotypic plasticity has resulted in consideration of the role of maternal effects⁷. In a variable environment, mothers can transmit information about environmental variability to the next generation, and influence the adaptive phenotypic response of their offspring⁸. For example, in *Daphnia*, such maternal effects have recently been reported for induced defences in response to predator-borne chemicals⁹. Maternal control of diapause has been reported for insects and mites⁷; the main factors in the maternal environment triggering the developmental switch in larvae are photoperiod and temperature¹⁰.

In *Daphnia*, bisexual reproduction and resting-egg formation is supposed to be induced by changing environmental conditions^{1,2} (that is, a phenotypic response) so individuals can alternate between modes of reproduction. Multiple environmental cues are required to induce resting-egg production, among them food availability, photoperiod and population density^{1,2}. Besides maternal nutrition³, another factor where transgenerational information can be expected is photoperiod, as the day length experienced by the mother and the day length experienced by succeeding generations can inform the offspring about the season in which they live¹⁰. With respect to the switch to resting-egg production, knowledge about the season is particularly important for *Daphnia*. As they are short-lived (one week from birth to maturation), measurements of the change in day length spread over two generations would be more reliable. Therefore, we raised *Daphnia pulicaria* in an experiment under all possible combinations of food conditions and photoperiod in the maternal and offspring generations, and measured the readiness of offspring to produce resting eggs (see Methods for details). Even at peak times of resting-egg production, only a few per cent of females in a natural population carry ehippia^{11,12}. In order to increase the incidence of ehippia in the experiments, we treated all offspring with a 'shock'—a complex of diapause-inducing factors—as soon as they matured (see Methods).

None of the daphnid offspring produced resting eggs in a high-food environment. This is only partly consistent with recent experimental results³ that suggested that ehippia production in *Daphnia* requires an asymmetrical mismatch in food conditions between mother and offspring: that is, a 'step-down' change in food. Using multiple cues, we could induce resting eggs when both mother and offspring lived at low food conditions, but never with

Table 1 Effect of environment on resting-egg production

Maternal environment	Offspring environment			
	High food		Low food	
	SD	LD	SD	LD
High food				
SD	0	0	52.3 (7.4)	38.0 (14.7)
LD	0	0	13.0 (7.2)	11.0 (11.0)
Low food				
SD	0	0	7.5 (7.5)	15.8 (4.6)
LD	0	0	0	30.7 (15.2)

Effect of maternal and offspring environment on the mean ($\pm$ s.e.) percentage of daphnids producing resting eggs (ephippia) within two clutches after the environmental shock (see Methods). Females and offspring experienced either high (1 mg C l⁻¹) or low (0.2 mg C l⁻¹) food concentrations in combination with short (10 h light; SD) or long (16 h light; LD) photoperiod. Kruskal–Wallis ANOVA by ranks yielded a highly significant effect of offspring food concentration ($H = 22.3$, $n = 48$, $P < 0.0001$). Pairwise comparisons (Kruskal–Wallis test) within the offspring low-food treatments yield significant effects of mother's photoperiod for the maternal high-food treatments ($H = 5.85$, $n = 12$, $P = 0.016$) and of offspring photoperiod for the maternal low-food treatments ($H = 6.22$, $n = 12$, $P = 0.013$).

* Present address: Zoological Institute of the Russian Academy of Sciences, St Petersburg, 199034, Russia.

a 'step-up' approach (Table 1). Overall, offspring photoperiod did not have a significant effect.

Within the offspring low-food environment, we found a significant effect of maternal photoperiod, depending on the mother's food environment. If the mother lived under good food conditions, the photoperiod she had experienced affected the resting-egg production of offspring. Offspring of short-day mothers produced significantly more resting eggs than the offspring of long-day mothers (Table 1). This is a transgenerational effect: that is, the mother transmits information about photoperiod to her offspring through the eggs. The pattern looks different if the mother lived under poor food conditions, which means she had less energy available to provision the eggs. There is no significant effect of the mother's photoperiod, but offspring photoperiod has an effect, which results in the striking difference between ephippia numbers in the maternal long-day treatment. Although there is no direct maternal effect of photoperiod, the differing ephippia production is notable, as the mother's photoperiod determines the response of offspring to their photoenvironment. If the mother lived under short days, the photoperiod experienced by offspring has no effect on resting-egg production; offspring produce moderate numbers of ephippia in any case. However, resting-egg production by offspring is either suppressed (at short days) or enhanced (at long days) if the mother lived in long-day conditions. We have no information about the nature of the transgenerational signal. The strong effect of food suggests that the mother's energy content has a role, but it is also possible that a chemical signal is involved, as found for diapause induction in flesh fly pupae¹³.

Although this study was originally performed with a physiological focus, we think that the results may be interpreted from an ecological point of view. Such an interpretation is a little hampered by the shock treatment, which provided a third stimulus that is not normally found so strongly in nature. However, the shock was applied to all treatments equally and only after the offspring had already reached maturity, thus it could not interfere with the maternal information. In fact, it did not induce the formation of a single ephippium in the offspring high-food treatments (Table 1). We believe that, despite the somewhat artificial experimental design, our results may explain the seasonal timing of resting-egg production by *Daphnia* in temperate lakes, as discussed below.

Typically¹⁴, temperate lakes produce two seasonal maxima of edible algae in spring (April/May) and summer separated by a low-food period (clear-water phase), which results from heavy grazing

by abundant *Daphnia*. In autumn, the algal population vanishes quickly due to unfavourable abiotic conditions (Fig. 1). This creates differing food–photoperiod combinations. The autumn situation is represented by a high-food/short-day environment for the mother, and a low-food/short-day environment for the offspring, which yielded the highest response in our experiments. The readiness to respond was much lower if the mother had already been under low-food conditions. Days in spring (May/June) are longer than in late autumn, but day length is still increasing, thus offspring experience slightly longer days than mothers do. With deteriorating food conditions after the spring algal maximum, daphnids are ready to respond with a first burst of ephippia production. Even the mother's poor-food environment in early summer does not prevent offspring from reacting with moderate numbers of ephippia produced under long-day conditions. Insects have been found to respond to absolute day length or changes in photoperiod^{10,15}. In our experiments, *D. pulicaria* responded with resting-egg production to both constant and differing transgenerational photoperiods, but some of the photoperiod–food combinations may not be relevant to the field: for example, a low-food environment for mother and daughter at decreasing day length. The data in Fig. 1 suggests that the absolute day length is more important than the relative change.

Placing our experimental data in the field context results in a reasonable pattern. Pulses of ephippia production are often observed after the algal peaks. In coincidence with our experimental results (Fig. 1), the strongest ephippia production is mostly observed in late autumn. These resting eggs, having survived the harsh winter conditions in diapause, provide hatchlings for the start of the population in the next spring, in addition to a few overwintering parthenogenetic females. This general pattern holds, although different species may contribute to the two peaks in varying proportions, and even clones of the same species react differently^{11,12}. Our laboratory data predict the two peaks, thus providing evidence that the transgenerational effect is important. In addition to the environmental stimulus, transgenerational effects may help to fine-tune the response. The correct timing of ephippia production must be under strong natural selection, as diapause is an essential component of the life-cycle in *Daphnia* and many other animals and plants. The transmission of day-length information to the next generation may provide an important clue about the season in short-lived organisms. □

Methods

We used a single clone of *D. pulicaria* Forbes that was originally isolated from a fish pond in southern Germany more than 20 years ago. This clone is particularly suitable for experiments on diapause induction as it is obligately parthenogenetic: that is, it can produce resting eggs without fertilization by males, a phenomenon described particularly for holarctic strains of *Daphnia*¹⁶. All daphnids were cultivated at 20 °C in 250-ml flow-through vessels (8–12 females per vessel). The flow-through system assures a constant food environment for the filter-feeding *Daphnia* at low food concentrations¹⁷. Green algae—*Scenedesmus obliquus*, a good food for *Daphnia*—were offered at concentrations of 0.2 mg Cl⁻¹ (low food) or 1.0 mg Cl⁻¹ (high food). Room light was adjusted to either 10 h d⁻¹ (short day) or 16 h d⁻¹ (long day). Before the *Daphnia* clone was used for the experiment, it had been kept at dim, constant light for several years, thereby excluding photoperiod effects.

An experiment began with growing neonates of *D. pulicaria* in four combinations of high and low food and short and long day until they reached maturity after 6–10 d, depending on food availability (maternal environment). They were kept under these conditions for the next two moults. Before these females laid their third clutch of eggs, they were transferred to 16 combinations of maternal and offspring environment with respect to food and photoperiod. This ensured that eggs in the brood pouch did not experience the mother's photoperiod. Each treatment combination was run in triplicate. Neonates were then raised under the new conditions (offspring environment) until they themselves reached maturity. As soon as the offspring *Daphnia* had produced their first clutch of subitaneous eggs, they were treated with an environmental 'shock' that is supposed to enhance the formation of resting eggs (24 h of starvation followed by two days at low food, 16 °C and 8 h day length). They were then transferred back into their original conditions, and monitored for the production of ephippia in the next two clutches. The proportion of ephippia-producing daphnids per vessel in two clutches was used as the response variable. Owing to inhomogeneous variances, results could not be analysed by general analysis of variance (ANOVA). We therefore used a non-parametric Kruskal–Wallis test.

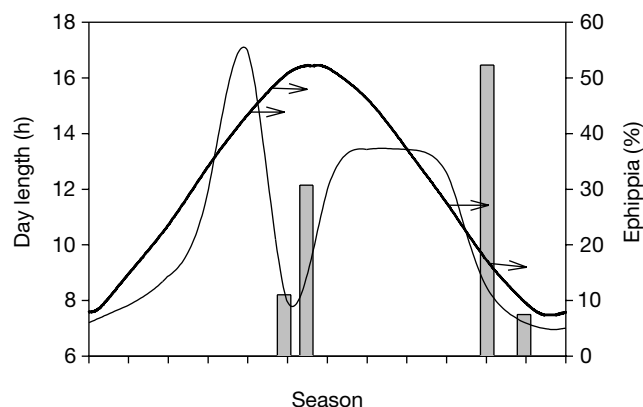


Figure 1 Schematic representation of the seasonal change of environmental factors triggering *Daphnia* resting-egg formation in a temperate, mesotrophic/eutrophic lake. Thick line, day length; thin line, algal biomass (relative units) according to the PEG model¹⁴. Bars represent the percentage of *Daphnia* that produced resting eggs in our laboratory experiments under the combinations of food and day length for mothers and offspring indicated by arrows. Our experiments predict two seasonal periods of resting-egg production.

Received 12 September; accepted 26 October 2001.

1. Stross, R. G. & Hill, J. C. Diapause induction in *Daphnia* requires two stimuli. *Science* **150**, 1463–1464 (1965).
2. Kleiven, O. T., Larsson, P. & Hobaek, A. Sexual reproduction in *Daphnia magna* requires three stimuli. *Oikos* **65**, 197–206 (1992).
3. LaMontagne, J. M. & McCauley, E. Maternal effects in *Daphnia*: what mothers are telling their offspring and do they listen? *Ecol. Lett.* **4**, 64–71 (2001).
4. Hairston, N. G. Jr, Van Brunt, R. A., Kearns, C. M. & Engstrom, D. R. Age and survivorship of diapausing eggs in a sediment egg bank. *Ecology* **76**, 1706–1711 (1995).
5. Hairston, N. G. Jr *et al.* Rapid evolution revealed by dormant eggs. *Nature* **401**, 446 (1999).
6. Cousyn, C. *et al.* Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proc. Natl Acad. Sci. USA* **98**, 6256–6260 (2001).
7. Mousseaux, T. A. & Fox, C. W. The adaptive significance of maternal effects. *Trends Ecol. Evol.* **13**, 403–407 (1998).
8. Fox, C. W. & Mousseaux, T. A. in *Maternal Effects as Adaptations* (eds Mousseaux, T. A. & Fox, C. W.) 159–177 (Oxford Univ. Press, Oxford, 1998).
9. Agrawal, A. A., Laforch, C. & Tollrian, R. Transgenerational induction of defences in animals and plants. *Nature* **401**, 60–63 (1999).
10. Danilevskii, A. A. *Photoperiodism and Seasonal Development in Insects* (Oliver & Boyd, Edinburgh, 1965).
11. Spaak, P. Sexual reproduction in *Daphnia*: interspecific differences in a hybrid species complex. *Oecologia* **104**, 501–507 (1995).
12. Caceres, C. A. Interspecific variation in the abundance, production and emergence of *Daphnia* diapausing eggs. *Ecology* **79**, 1699–1710 (1998).
13. Webb, M. L. Z. & Denzlinger, D. L. GABA and picrotoxin alter expression of a maternal effect that influences pupal diapause in the flesh fly, *Sarcophaga bullata*. *Physiol. Entomol.* **23**, 184–191 (1998).
14. Sommer, U., Gliwicz, Z. M., Lampert, W. & Duncan, A. The PEG-model of seasonal succession of planktonic events in fresh waters. *Arch. Hydrobiol.* **106**, 433–471 (1986).
15. Tauber, M. C., Tauber, C. A. & Masaki, S. *Seasonal Adaptations of Insects* (Oxford Univ. Press, New York, 1986).
16. Hebert, P. D. N. & Crease, T. Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. *Heredity* **51**, 353–369 (1983).
17. Lampert, W., Schmitt, R.-D. & Muck, P. Vertical migration of freshwater zooplankton: test of some hypotheses predicting a metabolic advantage. *Bull. Mar. Sci.* **43**, 620–640 (1988).

Acknowledgements

We thank M. Volquardsen for technical support, and C. Jamieson for linguistic help. V.A. was supported by a fellowship of the Max Planck Society and a grant of the Russian Science Foundation for Basic Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.A. (e-mail: alekseev@va4186.spb.ru) or W.L. (e-mail: lampert@mpil-ploen.mpg.de).

The genetic architecture of divergence between threespine stickleback species

Catherine L. Peichel*, Kirsten S. Nereng*, Kenneth A. Ohgi*, Bonnie L. E. Cole*, Pamela F. Colosimo*, C. Alex Buerkle†, Dolph Schluter‡ & David M. Kingsley*

* Department of Developmental Biology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305-5329, USA

† Department of Biology, University of Wisconsin-Eau Claire, Eau Claire, Wisconsin 54702-4004, USA

‡ Zoology Department and Centre for Biodiversity, University of British Columbia, Vancouver, British Columbia V6T 1ZT, Canada

The genetic and molecular basis of morphological evolution is poorly understood, particularly in vertebrates. Genetic studies of the differences between naturally occurring vertebrate species have been limited by the expense and difficulty of raising large numbers of animals and the absence of molecular linkage maps for all but a handful of laboratory and domesticated animals. We have developed a genome-wide linkage map for the three-spined

stickleback (*Gasterosteus aculeatus*), an extensively studied teleost fish that has undergone rapid divergence and speciation since the melting of glaciers 15,000 years ago¹. Here we use this map to analyse the genetic basis of recently evolved changes in skeletal armour and feeding morphologies seen in the benthic and limnetic stickleback species from Priest Lake, British Columbia. Substantial alterations in spine length, armour plate number, and gill raker number are controlled by genetic factors that map to independent chromosome regions. Further study of these regions will help to define the number and type of genetic changes that underlie morphological diversification during vertebrate evolution.

Three-spined sticklebacks provide one of the best known examples of rapid adaptive radiation in vertebrates. A large number of distinct morphological forms of sticklebacks have evolved following the colonization of newly created coastal streams and lakes at the end of the last ice age¹. In at least six lakes in coastal British Columbia, pairs of sympatric stickleback species have been identified. Members of a species pair are adapted to different niches within a lake, with corresponding changes in feeding morphology and defensive armour occurring in parallel in the different lakes (Fig. 1)². The benthic species feeds on invertebrates near shore and has a great reduction in the amount of body armour, increased body depth, and a decreased number of gill rakers for filtering ingested food. The limnetic species more closely resembles an ancestral marine fish, with more extensive body armour, a longer and more streamlined body, and an increased number of gill rakers. Despite reproductive isolation between the two species in the wild^{3–6}, it is possible to establish productive matings between the two species under laboratory conditions². The resulting F₁ hybrids are viable and fertile, making it possible to carry out a formal genetic analysis of the number and location of loci responsible for the adaptive morphological differences between these naturally occurring vertebrate species.

To develop resources for genome-wide linkage mapping in *Gasterosteus aculeatus*, we used large-scale library screening and sequencing to identify a collection of genomic and complementary DNA clones containing microsatellite repeat sequences. Initially, we sequenced 192 kilobases (kb) of random genomic clones and showed that CA dinucleotides were the most common form of microsatellite in sticklebacks, occurring approximately once every 14 kb. We subsequently screened genomic and cDNA libraries with a (GT)₁₅ probe, sequenced 3,560 clones, and identified 1,176 new microsatellite loci. Primers flanking 410 new and 18 previously identified microsatellites^{7–9} were designed and used to type a genetic cross between the benthic and limnetic species from Priest Lake,

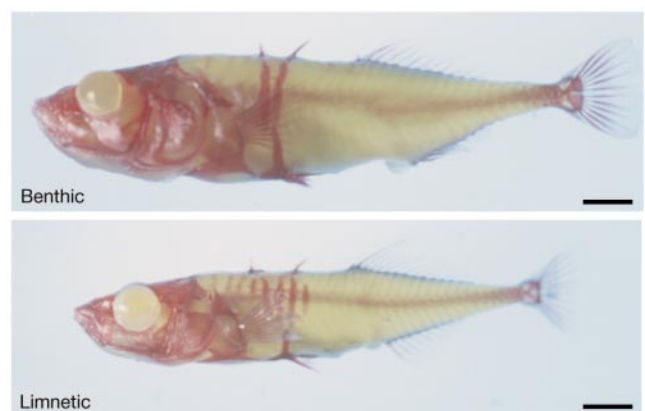


Figure 1 Representative benthic and limnetic fish from Priest Lake, British Columbia, are stained with alizarin red to highlight bone. The benthic fish are larger, more deep-bodied, and have fewer bony lateral plates than the limnetics. Scale bars, 5 mm.

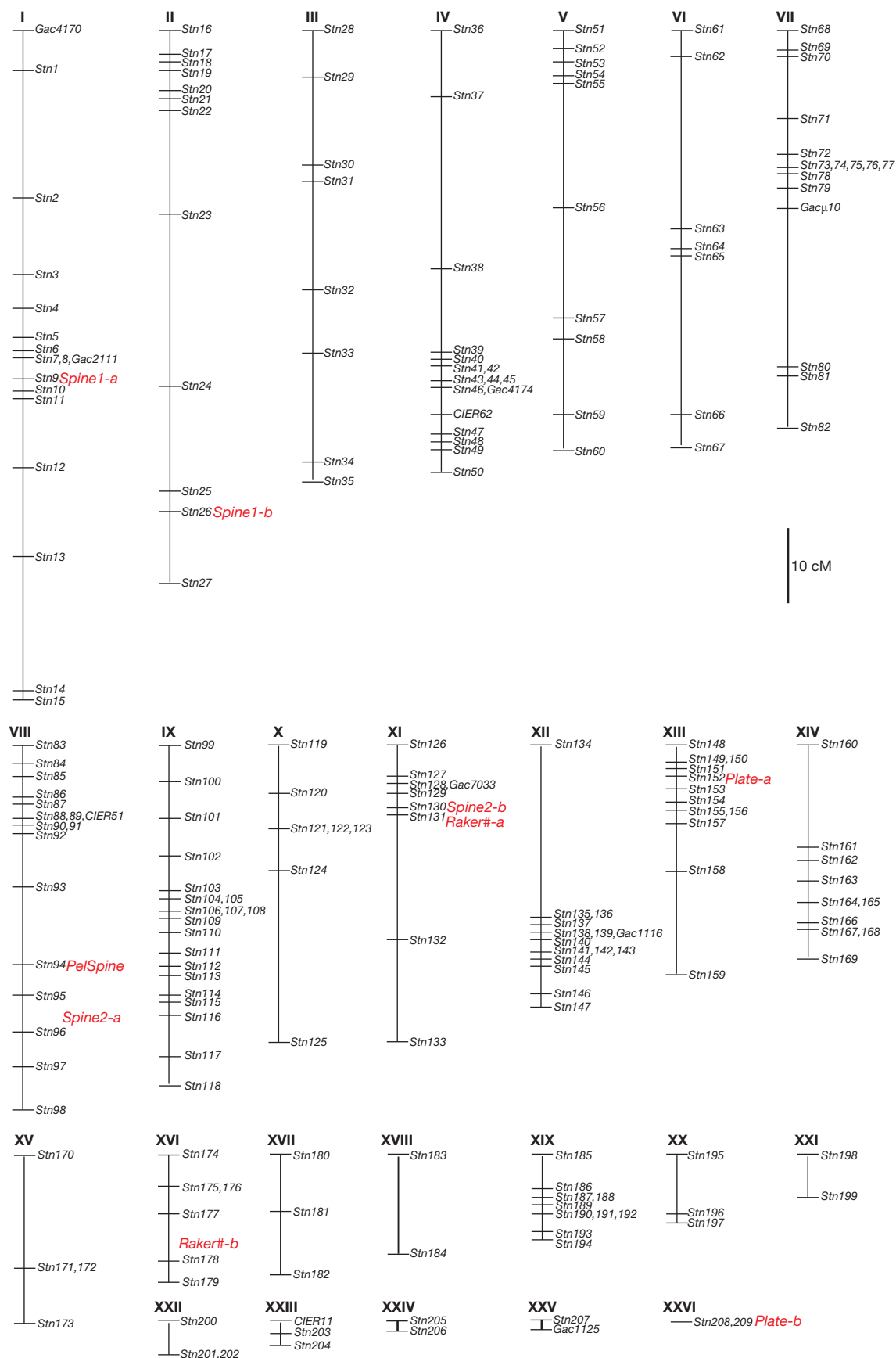


Figure 2 Genetic linkage map of *Gasterosteus aculeatus*. Each linkage group has been assigned a roman numeral (I to XXVI) in order of total genetic length. Microsatellite loci identified at Stanford are designated with a prefix of *Stn*, followed by a number which is

based on its serial position in the initial map. Twelve previously published microsatellites are also included on the map⁷⁻⁹. The map locations of quantitative trait loci (QTLs) affecting feeding morphology and skeletal armour are shown in red.

British Columbia (Fig. 1). For this cross, an individual Priest benthic female was mated with a single Priest limnetic male, and a single F_1 male (B_1L_1) was crossed to a second Priest benthic female (B_2B_3) to generate 103 progeny. Of the 281 markers that amplified robust bands from the F_1 and benthic parent, 227 (81%) were polymorphic, and therefore informative, in one or both parents. Higher rates of polymorphism were seen in the F_1 male than the benthic female parent (71% versus 57% of 281 markers), consistent with a greater level of genetic diversity between the distinct populations of benthic and limnetic fish than within the benthic population.

The segregation patterns of the 227 informative markers were scored on 92 progeny from the cross, and the 20,884 resulting genotypes were analysed for linkage using JoinMap software¹⁰. The markers were ordered into 26 linkage groups covering a total genetic distance of 886 centimorgans (cM), using a conservative LOD score (log likelihood ratio) threshold of 4.0 (Fig. 2). *Gasterosteus aculeatus* has a total of 21 chromosomes¹¹; therefore, we expect that some current linkage groups will collapse with other groups as additional markers are added to the map. Over 96% of the markers were linked to other markers on the map with an average density of 1 marker per 4 cM, suggesting a high probability that the existing markers can be used for genome-wide linkage mapping of interesting traits in many different stickleback populations.

Previous studies have shown benthic and limnetic species have distinct trophic morphologies adapted to feeding on either small invertebrates in the near shore environment, or zooplankton in the open water^{12,13}. Limnetic fish have larger eyes, longer snouts and jaws, and more numerous gill rakers, which are morphological adaptations that enhance feeding performance on small zooplankton¹². To examine the influence of different genetic regions

on trophic morphology, we counted the number of both long and short gill rakers on the first gill arch of all progeny from the Priest Lake cross (Fig. 3b). No major quantitative trait loci (QTLs) were found that influence the number of long gill rakers in the cross consistent with previous biometrical studies suggesting that gill raker number may be based on a large number of genes of small effect^{14,15}. In contrast, the number of short gill rakers is influenced by two QTLs that map to separate linkage groups. Together, these two QTLs accounted for nearly two-thirds of the variance in small raker number (Table 1, Figs 2 and 3c). Evolutionary change in the number of short gill rakers may thus be influenced by genetic effects at a relatively small number of chromosome regions.

The amount of skeletal armour is one of the most striking morphological differences between different stickleback populations, including the benthic and limnetic species pairs. Benthics have reduced armour, often with a reduced or absent first dorsal spine, reduced or absent pelvic spines, and a reduced number of lateral plates (Fig. 1). These changes in skeletal armour may be related to the different predation regimes experienced in the near-shore and open-water environments. Open-water populations experience more bird and fish predation, where longer dorsal and pelvic spines appear to offer greater protection¹⁵⁻¹⁷. In contrast, near-shore populations experience predation by insects, an environment in which spine reduction may be advantageous, and may be accompanied by loss of the lateral plates that support the dorsal and pelvic spines^{17,18}.

Linkage analysis of spine lengths and lateral plate number in the Priest cross identified QTLs that influence the length of the first and second dorsal spine, the pelvic spine, and the number of lateral plates (Table 1, Figs 2 and 3a, e). These QTLs accounted for 17 to

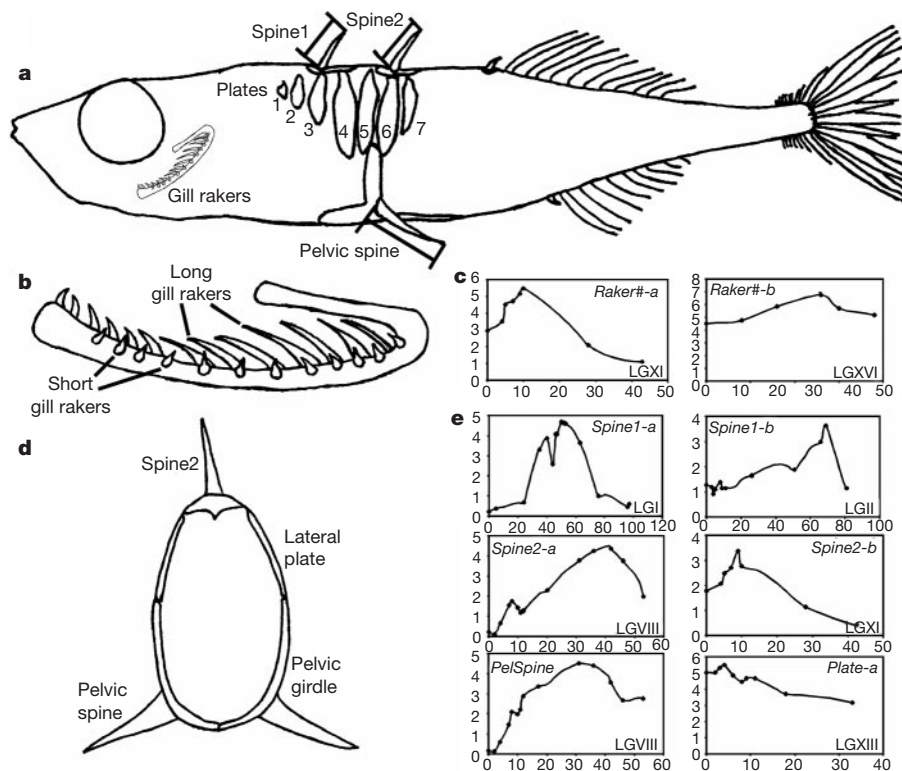


Figure 3 Mapping of morphological traits in the Priest Lake cross. Morphological measurements and traits for which significant QTLs were found: **a**, **d**, Skeletal armour shown in profile (**a**) or cross-section (**d**, adapted from ref. 30); **b**, gill raker numbers. The LOD scores (y-axis) were graphed relative to position in cM along the linkage group (x-axis). **c**, Number of small gill rakers; **e**, armour traits. Dots indicate the LOD score for markers on

the linkage group. The lines were drawn by plotting the LOD scores calculated by MapQTL at each marker as well as at 5.0-cM intervals along the linkage group. A graph is not shown for the *Plate-b* locus on linkage group XXVI, which consists of two non-recombinant markers.

26% of the total variance in each trait, and mapped to several distinct linkage groups. The locations of the QTLs influencing the first and second dorsal spines were completely distinct, suggesting that very similar morphological features can be influenced by different genetic regions. In contrast, the length of the second dorsal spine and the pelvic spine were both influenced by QTLs that map to a similar region of linkage group VIII. These two spines define the maximal dorsal and ventral extent of stickleback armour (Fig. 3d) and are thought to have an important role in defence against gape-limited predators^{15–17}. Variation in the length of these functionally related spines may be due to pleiotropic effects of the same locus or to linked genetic factors on linkage group VIII. Co-localization of ecologically important QTLs has recently been found in other systems, and may influence patterns of naturally occurring variation, adaptation and speciation^{19,20}.

The variation in length of both the second dorsal spine and the pelvic spine was due primarily to allelic differences segregating within the benthic population (Table 1; see also Supplementary Information). Previous studies have found significant polymorphism for morphological traits within stickleback populations²¹, perhaps maintained by spatial or temporal variation in particular habitats, or contrasting advantages of different skeletal phenotypes at different stages of the stickleback life cycle^{22,23}. In contrast, the length of the first dorsal spine and number of lateral plates and short gill rakers were influenced by allelic differences both between and within populations (Table 1 and Supplementary Information). Although the presence of a limnetic allele was usually associated with an increase in length or number of skeletal elements, this effect was sometimes seen only in the presence of a particular benthic allele. For example, a limnetic allele at the *Plate-a* QTL was associated with higher mean lateral plate number in combination with the B₃ benthic allele, but not the B₂ benthic allele, suggesting that allelic interactions influence this phenotype (Table 1 and Supplementary Information). For those traits affected by two distinct QTLs, the presence of limnetic alleles at both QTLs often caused a substantial phenotypic effect. For example, fish with 0, 1 or 2 limnetic alleles at the *Plate-a* and *Plate-b* QTL showed an increase from 7.9 to 9.0 to 10.5 in the mean number of lateral plates, and fish with 0, 1 or 2 limnetic alleles at the *Spine1-a* and *Spine1-b* QTL showed an increase from 0.12 mm to 1.0 mm to 1.42 mm in the mean length of the first dorsal spine (Supplementary Information). Therefore, a 33% difference in armour plate number and a greater than tenfold increase in the size of the first dorsal spine can be influenced by genetic effects at a

relatively small number of chromosome regions.

Our results in this vertebrate system are consistent with a number of recent genetic studies in plants and insects suggesting that evolutionary changes between organisms are controlled by genes with a variety of magnitudes of effect, some of which account for a substantial fraction of the variance in particular traits²⁴. The small size of the current cross limits our ability to detect additional genes of smaller effect, and could overestimate the contribution of minor QTL²⁵. Larger crosses are now needed to identify the full spectrum of genetic changes that contribute to the morphological differences between benthic and limnetic fish, and to narrow the location of the gene or genes within each chromosome interval that contribute to morphological divergence.

The many different chromosome regions that affect specific aspects of skeletal anatomy in sticklebacks reveal a flexible genetic system for independent modification of the size and number of different feeding and armour structures. However, some functionally related traits map to similar chromosome regions, suggesting that genetic linkage or pleiotropy may help account for the covariation in dorsal and pelvic spine lengths previously reported in many stickleback populations¹⁵. An extensive literature already exists on the great variation in size, morphology, colour and behaviour of freshwater sticklebacks around the world¹. The ease of collecting and crossing such fish in the laboratory, the relatively compact genome size of sticklebacks²⁶, the development of genome-wide linkage maps, and the ability to detect QTL for important evolutionary differences provide a particularly favourable system for further molecular studies of the genetic basis of morphological and behavioural changes during vertebrate evolution. □

Methods

Library construction, screening and sequencing

Genomic DNA from a single Paxton lake adult fish was digested with *RsaI* and fragments from 0.8–1.6 kb were cloned into the *EcoRV* site of pBluescriptSK(+). An oligo(dT) primed cDNA library was constructed in lambda ZAP Express (Stratagene) using RNA isolated from the head and internal organs of two adult fish from Salinas River, California. Colony and phage lifts were hybridized with an end-labelled oligonucleotide (GT)₁₅ in 4 × SSPE, 1 × Denhardt's solution, and 1% SDS at 60 °C overnight, and then washed for 2 × 10 min in 0.1 × SSC and 0.1% SDS at 60 °C. All positive clones were grown overnight in 96-well plates, and plasmid DNA was prepared using the QIAprep96 Turbo miniprep (Qiagen). Each clone was sequenced with a T3 and a T7 primer on a 96-lane gel on an ABI377 sequencer. Clones containing a microsatellite repeat were analysed with Primer3 (ref. 27) to identify primer pairs flanking the microsatellite. Primer pairs were picked to have melting temperatures between 55–65 °C and to give products between 100–250 bp. Forward primer pairs were labelled with one of three 5' phosphoramidite fluorescent conjugates: 6-fam (blue), 6-hex (yellow) or tet (green).

Table 1 Location and magnitude of effect for QTLs

Trait	Locus name	Linkage group	LOD	PVE (%)	Phenotype means			
					L ₁ B ₂	B ₁ B ₂	L ₁ B ₃	B ₁ B ₃
Feeding modifications								
Gill raker numbers	<i>Raker#-a</i>	XI	5.5*	26	15.4	14.3	13.8	13.8 rakers†‡§
	<i>Raker#-b</i>	XVI	6.8*	37	14.9	13.2	14.4	14.3 rakers†§
Body armour								
Lateral plate number	<i>Plate-a</i>	XIII	5.5*	26	8.5	8.6	10.6	8.4 plates†‡§
	<i>Plate-b</i>	XXVI	4.6*	22	10.0	8.2	9.6	8.0 plates†
Dorsal spine 1 length	<i>Spine1-a</i>	I	4.7*	21	1.53	0.49	1.00	0.57 mm†
	<i>Spine1-b</i>	II	3.6	17	1.42	0.73	0.97	0.13 mm†‡
Dorsal spine 2 length	<i>Spine2-a</i>	VIII	4.5*	22	2.72	2.59	2.31	2.35 mm‡
	<i>Spine2-b</i>	XI	3.4	17	2.57	2.65	2.22	2.39 mm‡
Pelvic spine length	<i>PelSpine</i>	VIII	4.5*	25	3.35	3.41	2.91	2.84 mm‡

For each QTL detected, the linkage group maximum LOD score, and percentage of the phenotypic variance explained (PVE) are indicated. QTLs were scored as significant or suggestive based on conservative recommendations for genome-wide linkage mapping²⁸, and by permutation testing for each trait. The genome-wide LOD significance thresholds are 4.2 for gill raker number and pelvic spine length, and 4.3 for plate number and dorsal spine length. The chromosome-wide LOD suggestive thresholds are 2.9 for length of dorsal spine 1 on linkage group II, and 2.6 for length of dorsal spine 2 on linkage group XI. Mean phenotypic values of each trait were also calculated for those progeny that inherited either all benthic alleles (B₂B₂ and B₃B₃), or both limnetic and benthic alleles (L₁B₂ and L₁B₃) at the most closely linked microsatellite. Significant differences between phenotype means are the estimated effect of alternate alleles inherited from the F₁ male parent (L₁ or B₁, the between-species effect), alleles inherited from the benthic female parent (B₂ or B₃, the within-species effect) and/or their interaction. See Supplementary Information for details.

* Significant QTLs.

† Between-species effect.

‡ Within-species effect.

§ Interaction between alleles from L₁ or B₁ and B₂ or B₃.

Genotyping

All polymerase chain reactions were carried out in a PTC-200 DNA Engine thermocycler (MJ Research) in 10- μ l reactions containing 0.5 μ M of each primer, 5 ng DNA, 0.25 mM dNTPs (Pharmacia), 1.5 mM MgCl₂, and 0.25 units Taq polymerase (PE Applied Biosystems). The cycling conditions for all primer pairs were 1 cycle of 95 °C for 1 min 45 s, 56 °C for 45 s and 72 °C for 45 s; 5 cycles of 94 °C for 45 s, 56 °C for 45 s and 72 °C for 45 s; and 30 cycles of 90 °C for 45 s, 56 °C for 45 s and 72 °C for 45 s, followed by a final cycle of 72 °C for 5 min. PCR products from 3 to 6 different primer pairs were then pooled and analysed on a 96-lane gel on an ABI377 with Gene Scan 2.1 software (Applied Biosystems) and GENESCAN-500 TAMRA (Applied Biosystems) used as internal size standard.

Linkage map construction

A genetic linkage map was created using JoinMap version 2.0 (ref. 10) on a locus file containing genotypes of 227 microsatellite loci in 92 backcross progeny, with the population type set for segregation of up to four alleles per locus (cross-pollinator). The JMGRP module of JoinMap was used with a LOD score threshold of 4.0 to assign 219 of the 227 loci to 26 linkage groups. The JMREC module was then used on each of the linkage groups to determine phase information for each locus. For each linkage group, a map was created with the JMMAP module: Kosambi mapping function, LOD threshold of 0.001, recombination threshold of 0.499, jump threshold of 5.0, triplet value of 5.0, and no fixed order. A ripple was performed after all markers on the linkage group were added to the map.

Morphological analysis

Fish were fixed in 10% buffered formalin for at least 1 week, placed in distilled H₂O (dH₂O) for 24 h, stained with 0.008% alizarin red in 1% KOH for 24 h, placed in dH₂O for 24 h, and placed in 37% isopropyl alcohol for final storage. Measurements were done with Vernier callipers accurate to 0.02 mm. Lateral plates were counted on both sides of the body, and the number of long and short gill rakers were counted on the left side of the first gill arch.

QTL mapping

All morphological traits were analysed with MapQTL 3.0 (ref. 28) using the interval mapping method which fits a single QTL model that is based on four possible segregating genotypes and does not assume a particular model of relationship between benthic and limnetic alleles. The parameters used were: mapping step size of 5.0, maximum of 200 iterations, and functional tolerance value of 1.0e⁻⁸. A maximum of five flanking markers were used to resolve incomplete genotypes. Significance thresholds for linkage were chosen using conservative criteria for genome-wide linkage mapping in non-inbred individuals: suggestive linkage of LOD $\geq$ 3.2, significant linkage of LOD $\geq$ 4.5 (ref. 29). Significance thresholds were confirmed by permutation tests in MapQTL 4.0, with a genome-wide significance level of $\alpha = 0.05$, $n = 1,000$ for significant linkages and a chromosome-wide significance level of $\alpha = 0.05$, $n = 1,000$ for suggestive linkages. Suggestive loci were only reported for those traits which were also influenced by one or more significant QTL. Multiple QTL model (MQM) mapping with initial QTL did not change the results. Calculation of the percentage of phenotypic variance explained by a QTL was done in MapQTL 3.0 on the basis of the population variance found within the progeny of the cross.

Received 22 June; accepted 7 November 2001.

- Bell, M. A. & Foster, S. A. *The Evolutionary Biology of the Threespine Stickleback* (Oxford Science, New York, 1994).
- McPhail, J. D. in *The Evolutionary Biology of the Threespine Stickleback* (eds Bell, M. A. & Foster, S. A.) 399–437 (Oxford Univ. Press, New York, 1994).
- Ridgway, M. S. & McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): mate choice and reproductive isolation in the Enos Lake species pair. *Can. J. Zool.* **62**, 1813–1818 (1984).
- Nagel, L. & Schluter, D. Body size, natural selection, and speciation in sticklebacks. *Evolution* **52**, 209–218 (1998).
- Hatfield, T. & Schluter, D. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* **53**, 866–873 (1999).
- Vamosi, S. M. & Schluter, D. Sexual selection against hybrids between sympatric stickleback species: evidence from a field experiment. *Evolution* **53**, 874–879 (1999).
- Rico, C., Zadworny, D., Kuhnlein, U. & Fitzgerald, G. J. Characterization of hypervariable microsatellite loci in the threespine stickleback *Gasterosteus aculeatus*. *Mol. Ecol.* **7**, 271–272 (1993).
- Taylor, E. B. Microsatellites isolated from the threespine stickleback *Gasterosteus aculeatus*. *Mol. Ecol.* **7**, 925–931 (1998).
- Largiader, C. R., Fries, V., Kobler, B. & Bakker, C. M. Isolation and characterization of microsatellite loci from the three-spined stickleback (*Gasterosteus aculeatus* L.). *Mol. Ecol.* **8**, 342–344 (1999).
- Stam, P. & Van Ooijen, J. W. JoinMap, version 2.0: Software for the calculation of genetic linkage maps. (Centre for Plant Breeding and Reproduction Research, Wageningen, 1995).
- Chen, T.-R. & Reisman, H. M. A comparative study of the North American species of sticklebacks (Teleostei: Gasterosteidae). *Cytogenetics* **9**, 321–332 (1970).
- Bentzen, P. & McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): specialization for alternative trophic niches in the Enos Lake species pair. *Can. J. Zool.* **62**, 2280–2286 (1984).
- Schluter, D. Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. *Ecology* **74**, 699–709 (1993).
- Hatfield, T. Genetic divergence in adaptive characters between sympatric species of stickleback. *Am. Nat.* **149**, 1009–1029 (1997).

- Hagen, D. W. & Gilbertson, L. G. Geographic variation and environmental selection in *Gasterosteus aculeatus* L. in the Pacific Northwest, America. *Evolution* **26**, 32–51 (1972).
- Moodie, G. E. E. Predation, natural selection and adaptation in an unusual threespine stickleback. *Heredity* **28**, 155–167 (1972).
- Reimchen, T. E. Spine deficiency and polymorphism in a population of (*Gasterosteus aculeatus*): an adaptation to predators? *Can. J. Zool.* **58**, 1232–1244 (1980).
- Reimchen, T. E. Structural relationship between spines and lateral plates in threespine stickleback (*Gasterosteus aculeatus*). *Evolution* **37**, 931–946 (1983).
- Hawthorne, D. J. & Via, S. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412**, 904–907 (2001).
- Lynch, M. & Walsh, J. B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Massachusetts, 1998).
- Ahn, D. & Gibson, G. Axial variation in the threespine stickleback: genetic and environmental factors. *Evol. Dev.* **1**, 100–112 (1999).
- Reimchen, T. E. Predator-induced cyclical changes in lateral plate frequencies of *Gasterosteus*. *Behavior* **132**, 1079–1094 (1995).
- Swain, D. P. Selective predation for vertebral phenotype in *Gasterosteus aculeatus*: reversal in the direction of selection at different larval sizes. *Evolution* **46**, 998–1013 (1992).
- Orr, H. A. The genetics of species differences. *Trends Ecol. Evol.* **16**, 343–350 (2001).
- Beavis, W. D. in *Molecular Dissection of Complex Traits* (ed. Paterson, A. H.) 145–162 (CRC Press, Boca Raton, 1998).
- Hinegardner, R. Evolution of cellular DNA content in teleost fishes. *Am. Nat.* **102**, 517–523 (1968).
- Rozen, S. & Skaletsky, H. J. Primer3; code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html (1997).
- VanOoijen, J. W. & Maliepard, C. MapQTL, version 3.0: Software for the calculation of QTL positions on genetic maps. (Centre for Plant Breeding and Reproduction Research, Wageningen, 1996).
- VanOoijen, J. W. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* **83**, 613–624 (1999).
- Bell, M. A. in *Evolutionary Genetics of Fishes* (ed. Turner, B. J.) 431–528 (Plenum, New York, 1984).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Anderson, R. Melzer, K. Tsui, C. Uhlick and R. Vega for technical assistance. D.S. was supported by a Natural Sciences and Engineering Research Council of Canada research grant. C.L.P. is a research associate and D.M.K. is an assistant investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M.K. (e-mail: kingsley@cmgm.stanford.edu). GenBank accession numbers for stickleback microsatellite markers are G72126–G72334.

Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys

Christopher C. Pack, Vladimir K. Berezovskii & Richard T. Born

Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

In order to see the world with high spatial acuity, an animal must sample the visual image with many detectors that restrict their analyses to extremely small regions of space. The visual cortex must then integrate the information from these localized receptive fields to obtain a more global picture of the surrounding environment. We studied this process in single neurons within the middle temporal visual area (MT) of macaques using stimuli that produced conflicting local and global information about stimulus motion. Neuronal responses in alert animals initially reflected predominantly the ambiguous local motion features, but gradually converged to an unambiguous global representation. When the same animals were anaesthetized, the integration of local motion signals was markedly impaired even though neuronal responses remained vigorous and directional tuning characteristics were intact. Our results suggest that anaesthesia preferentially

affects the visual processing responsible for integrating local signals into a global visual representation.

Figure 1a shows an example of the 'aperture problem'^{1,2}: a moving edge viewed through a small aperture only provides information about the component of motion perpendicular to its orientation. In the primate visual system, the small receptive fields of neurons at stages up to and including the primary visual cortex (V1) function as apertures, and so are faced with this problem. Therefore, neurons in extrastriate cortex must integrate motion information in such a way as to correct for systematic errors in their inputs. In order to examine this process, we have recorded the responses of neurons in area MT—an important motion-processing region in extrastriate cortex—to stimuli designed to generate locally ambiguous motion measurements in V1 cells.

Figure 2a shows the response of a typical MT neuron in an alert monkey to the grating stimulus shown in Fig. 1a (see Methods for experimental details). The neuron responds strongly to motion to the left (its preferred direction), but weakly to other motion directions, and negligibly to rightward motion. Based on this tuning curve, we can make two different predictions about this neuron's response to the plaid stimulus in Fig. 1b. If the orientations of the component gratings of the plaid are significantly different, then when one grating moves to the left, the second grating will move in a direction that by itself generates no response. To the extent that the neuron 'sees' only the local motion cues generated by each grating, the tuning curve for the plaid stimulus should have a separate peak for each direction of plaid motion that moves one of the gratings leftward. The difference in these peaks should be close to the difference in the orientations of the gratings. This has been termed the 'component' prediction³. Alternatively, if the neuron integrates the global motion of the plaid stimulus, then the response should be strongest when the entire pattern moves leftward. This is the 'pattern' prediction³.

We tested this neuron with plaid stimuli consisting of two gratings that differed in their orientations by 144°. Figure 2b shows the earliest directionally selective response, which occurred at a latency of 62 ms. The direction tuning is clearly bimodal, consistent with the component prediction. However, the subsequent direction tuning, determined by averaging over the last 1,500 ms of stimulus presentation, is unimodal and peaks for leftward plaid motion (Fig. 2c), consistent with the pattern prediction. This indicates that the neural responses ultimately reflected the

computation of plaid pattern motion from the component gratings.

We note that human perception of the stimuli used in these experiments usually conforms to the pattern prediction^{4,5}, but the inputs from monkey V1 cells always conform to the component prediction^{3,6}. In this regard, we can say that the computation of pattern motion in MT is more closely correlated with normal perception. A simple way to test this idea is to repeat the experiment in the anaesthetized animal.

Figure 2d–f shows data from a different neuron recorded from approximately the same part of MT while the same animal was anaesthetized with isoflurane (see Methods for details). This neuron failed to compute the direction of plaid motion, responding instead to the plaid components for the duration of the stimulus presentation. We recorded 42 MT neurons in the anaesthetized animals (with monocular viewing), and 35 MT neurons while the same two animals were awake and fixating binocularly. We analysed the responses by computing the correlations³ between the observed

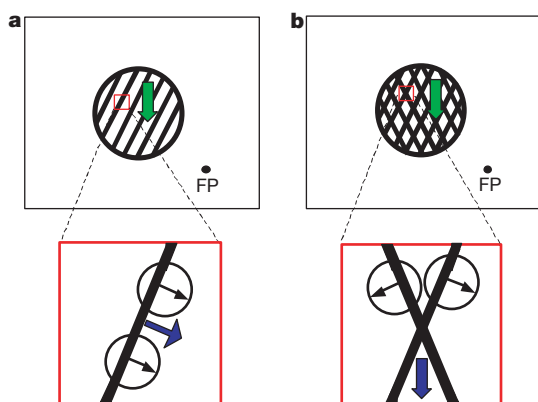


Figure 1 Experimental stimuli. Green arrows represent the actual direction of stimulus motion; black arrows, local motion cues; blue arrows, the perceived direction for human observers. Large circles in the top row depict an MT receptive field. **a**, A grating stimulus consisting of alternating black and white bars, the endpoints of which are not visible. **b**, A plaid stimulus constructed by superimposing two gratings. Because of the aperture problem, motion of the entire plaid pattern generates differing local motion cues (black arrows) that, if combined properly, yield the actual direction of plaid motion (blue arrow). Stimulus contrast has been reversed here for clarity. FP, fixation point.

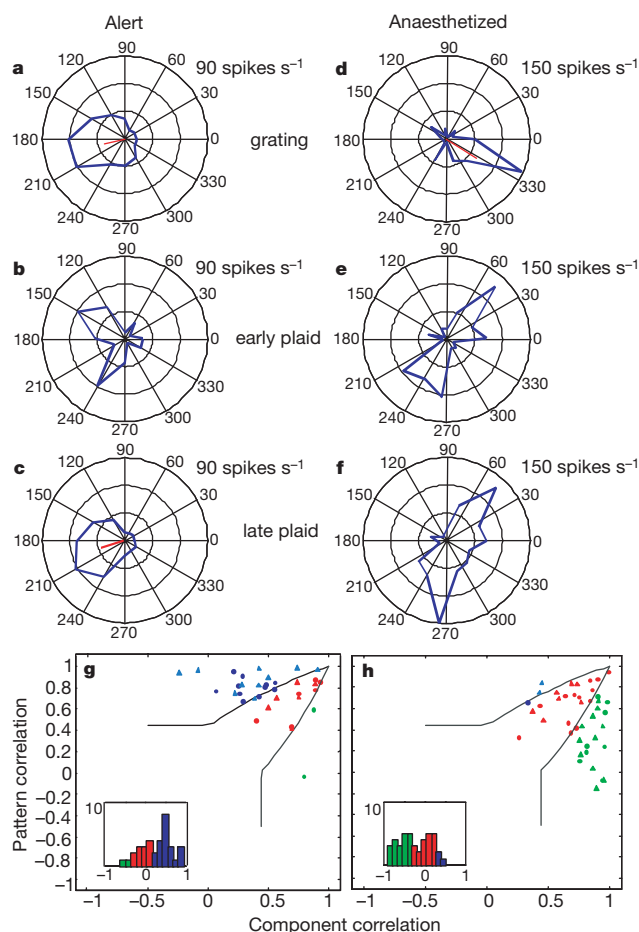


Figure 2 Results for plaid stimuli. **a**, Responses to grating motion for an MT cell recorded in an alert animal. Direction tuning is represented in polar coordinates, with the angle representing the grating direction and the radius representing the response in spikes per second. The red line indicates the vector average of the responses to each direction. **b**, Direction tuning of the first 20 ms of the cell's response to the plaid. **c**, Response of the same cell, averaged over the last 1,500 ms of stimulation. **d–f**, As for **a–c**, but in a different cell while the animal was anaesthetized. **g**, Classification of cells from alert animals according to correlations with the component and pattern predictions. Points upward and leftward from the black lines were classified as pattern cells (blue). Points downward and rightward from the black lines were classified as component cells (green). Points between the lines were poorly correlated with both predictions (red). The different symbols (triangle, circle) represent different monkeys. The histogram groups cells according to the strength of the pattern prediction. **h**, As for **g**, but cells recorded in anaesthetized animals.

plaid responses and the two predictions described above (component and pattern; see Methods). The results are shown in Fig. 2g and h. In the alert animals, 60% of the neurons were classified as pattern cells, compared with only 6% classified as pure component cells, based on the difference in the correlation coefficients. When the animals were anaesthetized, 7% were classified as pattern cells and 45% as component cells. This was not related to effects of anaesthesia on basic neuronal response properties, such as direction selectivity, nor to differences related to monocular versus binocular viewing (see below). A comparison of the tuning bandwidths for grating stimuli showed that the populations were not significantly different (anaesthetized mean 83.3° , alert mean 90.3° ; two-tailed t -test $P > 0.5573$), as has been shown previously for tuning curves in V1 (ref. 7) and MT⁸.

By subtracting the variance accounted for by the pattern prediction from that accounted for by the component prediction, we obtained a pattern index⁹, which describes the strength of the pattern prediction relative to the component prediction (see Methods). Positive values indicate close conformity with the pattern prediction, while negative values support the component prediction. Histograms of these values are shown in the lower left of Fig. 2g and h. A comparison of the pattern indices between the anaesthetized and alert populations indicates that these differences were highly significant (one-tailed t -test, $P < 0.0001$). Figure 3 shows that the shift from component responses to pattern responses in the alert animals (as in Fig. 2b, c) was typical of the MT population. The pattern index shifts from being strongly negative to strongly positive over a period of approximately 100 ms. This shift does not occur in the anaesthetized MT population. The same cells, when tested with tilted bar stimuli (as in ref. 10), responded primarily to the component of motion perpendicular to the orientation of the bars.

In order to ensure that these results were not due to differences between the two preparations (anaesthetized and alert) that were unrelated to anaesthesia, we conducted control experiments to test other potential confounding factors. Specifically, viewing was binocular in the alert case, but one eye was occluded in the anaesthetized experiments. Also, although the monkeys' vision was artificially corrected in the anaesthetized experiments (see Methods), some residual blurring may have occurred. To determine if these differences could explain our findings, we recorded from 10 additional neurons in an alert animal. For each neuron, three conditions were tested: binocular viewing, monocular viewing (one eye occluded), and monocular viewing with image blurred. The latter condition involved occluding one eye and defocusing the image by placing a +2-dioptre trial lens over the other eye. Of the 10

cells recorded, 5 were classified as pattern cells, and neither the monocular viewing nor the blurring caused any of these to switch to a component classification. The mean pattern indices in the binocular, monocular and blurred conditions were 0.20, 0.24 and 0.25. Pairwise comparisons of the component and pattern correlations across the different viewing conditions revealed no significant changes (two-tailed t -test following Fisher's z -transform, $P > 0.19$ in all cases). We conclude that it is extremely unlikely that these viewing differences were responsible for the differences in motion integration shown in Fig. 2.

Our results suggest that the process of resolving ambiguous local information into a global representation is significantly impaired by general anaesthesia. This explanation connects our recent findings in alert animals¹⁰ with previous neurophysiological studies in anaesthetized animals, which indicate that a minority of cells encode pattern motion^{3,11,12}. We also found that pattern motion is computed accurately by a minority of MT cells (7%) under anaesthesia, but that in alert animals this percentage increases greatly (60%). It should be noted that the percentage of MT pattern cells found in previous studies of anaesthetized animals^{11,12} has been closer to 25%, which is somewhat higher than what we found. This may be due to stimulus differences. For instance, our grating stimuli were separated by a larger orientation difference, and had a lower spatial frequency, than the gratings used in previous studies^{3,9}. The low spatial frequency was chosen to reduce the number of grating intersections, which increase in number with increasing spatial frequency¹³, and provide an unambiguous cue as to the actual pattern direction.

Computational models of motion integration generally consist of two or more stages. The first stage, usually assigned to V1, contains units that extract the component of motion perpendicular to the orientation of contours⁵. That is, they are limited by the aperture problem. Our finding that directional responses to grating stimuli are not impaired by anaesthesia suggests that this stage is largely intact in the anaesthetized animals. The second stage of motion processing, thought to correspond to MT, is responsible for integrating motion signals in such a way as to overcome the biases introduced in the first stage. It has been shown that this can be accomplished by an appropriate feedforward weighting of V1 inputs¹⁴. However, some elaboration of the feedforward model would be necessary to account for our observations on the temporal dynamics of motion integration.

One possible elaboration is that MT neurons receive information about the motion of features such as grating intersections via an intermediate 'non-Fourier' stage, requiring a slightly longer latency¹⁵. If this model is correct, our results suggest a selective

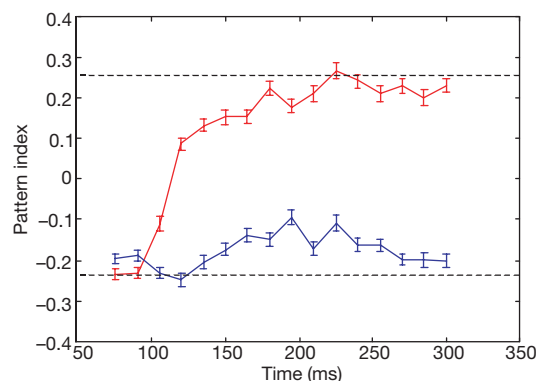


Figure 3 Change in the pattern index over time. Figure shows the mean pattern index for the populations of MT cells recorded from alert animals (red) and anaesthetized animals (blue). The horizontal dotted black lines indicate the mean value of the time-averaged

pattern index for the alert population (upper line) and anaesthetized population (lower line). Data are grouped in 15-ms bins. Error bars indicate standard error of the mean.

impairment of the intermediate stage. Such an effect might also explain the finding that MT neurons more accurately encode the direction of non-Fourier motion stimuli in alert animals¹⁶ than in anaesthetized animals¹⁷. Second, and more generally, anaesthesia could act on an early stage, by disrupting the temporal properties of neurons in the thalamus¹⁸. A third possibility is that the delayed response to the global properties of the stimulus, and its suppression by anaesthesia, are signatures of cortical feedback, as has been suggested by studies of contextual modulation in V1 (refs 19–21). Feedback could be useful for sharpening the tuning generated by feedforward inputs¹⁵ or for propagating unambiguous motion signals to ambiguous regions of the visual representation^{22,23}. Such mechanisms should be identifiable by manipulations such as cortical cooling, which have proven useful in the studies of functional connectivity in other parts of the visual system²⁴. □

Methods

Visual stimuli and recordings from alert monkeys

Recordings were obtained from single units in alert monkeys, as described previously²⁵. Each animal underwent a magnetic resonance imaging (MRI) scan to locate MT within the coordinates of a plastic grid inserted in the recording cylinder. The same grid, along with a guide tube, was used to guide insertion of the microelectrode. MT was identified on the basis of depth, prevalence of direction-selective neurons, receptive field size, and visual topography. Visual stimuli consisted of gratings and plaids, and were viewed binocularly at a distance of 57 cm. The gratings were square-wave, 0.5 cycles per degree with a duty cycle of 0.4. Plaids were constructed by overlapping two gratings moving in directions separated by 144°. There were no luminance differences at the intersections of the gratings. The resulting motion was informally judged to be completely coherent by human observers. Plaid speed was always faster than grating speed, as dictated by stimulus geometry⁵. All stimuli were generated by a PC containing a video board (SGT-Plus, Number Nine Corp), and had a luminance of 13.9 cd m⁻² against a dark background (0.025 cd m⁻²). For each cell, each stimulus was presented 5 times in blockwise random order for 2 s. To separate directional responses from on-transients and orientation responses, each stimulus remained stationary for 240 ms before moving. For gratings and plaids, motion direction was generally sampled at 24° intervals, although a few cells were measured at 30° intervals. Neuronal signals were recorded extracellularly using tungsten microelectrodes (FHC) with standard amplification and filtering (BAK Electronics), while the monkeys fixated a small spot. Fixation was monitored with an eye coil²⁶ and required to be within 1° of the spot for the monkeys to obtain a liquid reward. Single units were isolated using a dual time and amplitude window discriminator (BAK).

Recordings from anaesthetized monkeys

We used the same two monkeys for alert experiments and for recordings under general anaesthesia (isoflurane), as previously described²⁷. One monkey participated in two anaesthetized recording sessions, and the second monkey participated in one session. Alert recordings were obtained both before and after the anaesthetized recordings for both animals. To ensure proximity of the recording sites between the two preparations, electrodes were inserted into the same grid positions with the same guide tubes in the same recording chambers, restricting the variability in electrode position to 1 mm. This was confirmed by the nearly identical distributions of receptive field eccentricities in the alert and anaesthetized conditions (4°–18°). Anaesthesia was induced with ketamine (15 mg kg⁻¹, intramuscular). The animal was intubated, and anaesthesia was maintained with isoflurane (0.5–1% in oxygen). Electroencephalograms were not recorded. Throughout the experiment the electrocardiogram and end-tidal pCO₂ were continuously monitored, and any changes indicative of a lightening of anaesthesia were immediately corrected by increasing the level of isoflurane. The animal's head was immobilized in a stereotaxic frame using the same head post used in the alert experiments. The animal was paralysed with a continuous intravenous infusion of vecuronium bromide (0.05 mg kg⁻¹ h⁻¹). Gas-permeable contact lenses of the appropriate basal curvature were placed over the eyes. A streak retinoscope was used to measure the refractive state, and trial lenses were used to bring the stimulus display monitor, at a distance of 57 cm, into focus. The location of the fovea was back-projected onto the display monitor using an ophthalmoscope fitted with a reversing prism. This was re-checked periodically throughout the experiment.

Optics

In the alert animals, stimuli were viewed binocularly, whereas viewing in the anaesthetized experiments was monocular. Both monkeys had clear ocular media and were refracted while under anaesthesia with accommodation paralysed (1% atropine sulphate) using streak retinoscopy performed by an experienced operator (R.T.B.). Monkey 1 had a small hyperopic error of +0.25 dioptres and monkey 2 had a slightly larger hyperopic error (+1.75 dioptres). These were corrected using trial lenses during experiments under anaesthesia. For alert experiments, the stimuli were presented on a monitor that was 57 cm from the animal. This would require, for monkey 2, +3.5 dioptres of accommodation, which is well within the range of the accommodative mechanism of the primate eye.

Data analysis

For each cell, component and pattern predictions were generated from grating responses using an extension⁶ of the partial correlation measure³. Briefly, the component prediction is the sum of the observed grating responses rotated by 72° clockwise and anticlockwise. The pattern prediction is identical to the grating response. The partial correlation is given by $R_p = (r_p - r_c r_{pc}) / [(1 - r_c^2)(1 - r_{pc}^2)]^{1/2}$, where r_p is the raw correlation of the data with the pattern prediction, r_c is the raw correlation of the data with the component prediction, and r_{pc} is the correlation of the two predictions. The partial correlation coefficient for the component prediction, R_c , is obtained by exchanging r_c and r_p in the above equation. A measure of pattern selectivity relative to component selectivity is obtained by subtracting the variance accounted for by the component prediction from that accounted for by the pattern prediction. This pattern index is given by $(R_p^2 - R_c^2)$.

Received 30 August; accepted 31 October 2001.

- Wallach, H. Über visuell wahrgenommene Bewegungsrichtung. *Psychol. Forsch.* **20**, 325–380 (1935).
- Marr, D. & Ullman, S. Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B* **211**, 151–180 (1981).
- Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. The analysis of moving visual patterns. *Exp. Brain Res. Suppl.* **11**, 117–151 (1986).
- Stoner, G. R., Albright, T. D. & Ramachandran, V. S. Transparency and coherence in human motion perception. *Nature* **344**, 153–155 (1990).
- Adelson, E. H. & Movshon, J. A. Phenomenal coherence of moving visual patterns. *Nature* **300**, 523–525 (1982).
- Movshon, J. A. & Newsome, W. T. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* **16**, 7733–7741 (1996).
- Snodderly, D. M. & Gur, M. Organization of striate cortex of alert, trained monkeys (*Macaca fascicularis*): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. *J. Neurophysiol.* **74**, 2100–2125 (1996).
- Mikami, A., Newsome, W. T. & Wurtz, R. H. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *J. Neurophysiol.* **55**, 1328–1339 (1986).
- Stoner, G. R. & Albright, T. D. Neural correlates of perceptual motion coherence. *Nature* **358**, 412–414 (1992).
- Pack, C. C. & Born, R. T. Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* **409**, 1040–1042 (2001).
- Rodman, H. R. & Albright, T. D. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.* **75**, 53–64 (1989).
- Castel-Branco, M., Goebel, R., Neuenschwander, S. & Singer, W. Neural synchrony correlates with surface segregation rules. *Nature* **405**, 685–689 (2000).
- Alais, D., Wenderoth, P. & Burke, D. The size and number of plaid blobs mediate the misperception of type-II plaid direction. *Vision Res.* **37**, 143–150 (1997).
- Simoncelli, E. P. & Heeger, D. J. A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761 (1998).
- Wilson, H. R., Ferrera, V. P. & Yo, C. A psychophysically motivated model for two-dimensional motion perception. *Vis. Neurosci.* **9**, 79–97 (1992).
- Albright, T. D. Form-cue invariant motion processing in primate visual cortex. *Science* **255**, 1141–1143 (1992).
- O'Keefe, L. P. & Movshon, J. A. Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Vis. Neurosci.* **15**, 305–317 (1998).
- Mukherjee, P. & Kaplan, E. Dynamics of neurons in the cat lateral geniculate nucleus: in vivo electrophysiology and computational modeling. *J. Neurophysiol.* **74**, 1222–1243 (1995).
- Lamme, V. A. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15**, 1605–1615 (1995).
- Lamme, V. A., Zipser, K. & Spekreijse, H. Figure-ground activity in primary visual cortex is suppressed by anaesthesia. *Proc. Natl Acad. Sci. USA* **95**, 3263–3268 (1998).
- Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571–579 (2000).
- Hildreth, E. C. *The Measurement of Visual Motion* (MIT Press, Cambridge, Massachusetts, 1984).
- Lidén, L. H. & Pack, C. C. The role of terminators and occlusion cues in motion integration and segmentation: A neural network model. *Vision Res.* **39**, 3301–3320 (1999).
- Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
- Born, R. T., Groh, J. M., Zhao, R. & Lukasiewicz, S. J. Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron* **26**, 725–734 (2000).
- Robinson, D. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **10**, 137–145 (1963).
- Born, R. T. & Tootell, R. B. Spatial frequency tuning of single units in macaque supragranular striate cortex. *Proc. Natl Acad. Sci. USA* **88**, 7066–7070. (1991).

Acknowledgements

We thank P. Hendrickson for technical assistance, and M. Livingstone for comments on an earlier version of the manuscript. This work was supported by the NIH, The Whitehall Foundation and The Giovanni Armenise-Harvard Foundation for Scientific Research. C.C.P. was supported by a McDonnell-Pew fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.C.P. (e-mail: cpack@hms.harvard.edu).

Noggin and retinoic acid transform the identity of avian facial prominences

S.-H. Lee^{*†}, K. K. Fu^{*}, J. N. Hui^{*} & J. M. Richman^{*}

^{*} Department of Oral Health Science, Faculty of Dentistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

[†] Department of Dentistry, Medical College, Gyeongsang National University, Chinju, Korea

The signals that determine body part identity in vertebrate embryos are largely unknown, with some exceptions such as those for teeth and digits^{1,2}. The vertebrate face is derived from small buds of tissue, facial prominences, that surround the embryonic oral cavity³. In chicken embryos, the skeleton of the upper beak is derived from the frontonasal mass and maxillary prominences⁴. Here we show that bone morphogenetic proteins (Bmps) and the vitamin A derivative, retinoic acid (RA), are used to specify the identity of the frontonasal mass and maxillary prominences. Implanting two beads adjacent to the stage-15 presumptive maxillary field, one soaked in the Bmp antagonist Noggin⁵ and one soaked in RA, induces a duplicate set of frontonasal mass skeletal elements in place of maxillary bones. We also show that the duplicated beak is due to transformation of the maxillary prominence into a second frontonasal mass and not due to ectopic migration of cells or splitting of the normal frontonasal mass. Thus the levels of Bmp and RA determine whether specific regions of the face form maxillary or frontonasal mass derivatives.

Previous experiments^{6,7} in avian embryos have shown that the premigratory neural crest cells that populate the facial prominences contain some inherent skeletal patterning information. Thus, when premigratory neural crest cells are transplanted from midbrain to hindbrain, these heterotopic grafts cause beak-like outgrowths in the necks of host embryos^{6,7}. Neural crest cell migration ends by stage 14 (ref. 8) and then cells proliferate to fill the branchial arches and the frontonasal region (Fig. 1a). Immediately after migration is

Table 1 Frequency of duplicated structures

Type of treatment	Total N	Interorbital septum	Prenasal cartilage	Premaxilla	Egg tooth
Noggin, 1 mg ml ⁻¹	21	4	0	1	0
Noggin, high concentration*	14	5	2	4	5
Noggin + RA†	11	11	10	10	10
Noggin + RA 0.01 mg ml ⁻¹	5	1	3	1	1
Noggin + RA 1 mg ml ⁻¹ , beads left in	5	4	3	3	4

* Sequential soaking of beads as described in Methods.

† 1 mg ml⁻¹ Noggin + RA 0.1 mg ml⁻¹, both beads left in; or 1 mg ml⁻¹ Noggin + RA 1.0 mg ml⁻¹, RA bead removed at 6 h. These conditions were equivalent.

complete, first branchial arch cells can still change their fates¹, but this plasticity is gradually lost⁹. Within 24 h (stage 18) (ref. 10), the facial prominences (swellings of mesenchyme encased in epithelium) surround the oral cavity (Fig. 1b, c). Transplantation experiments demonstrate that each prominence gives rise to distinct skeletal elements and that by stage 20, the fate of each individual facial process is specified¹¹ (Fig. 1d). However, we do not know which molecules establish the identity of the facial prominences.

Bmps are good candidates for molecules that impart identity to each facial prominence. BMPs regulate expression of *Tbx* genes¹² that control limb identity¹³. At stage 15 when the facial prominences have not yet formed¹⁰, *Bmp* genes are expressed in the presumptive mesenchyme of the maxillary region between the first branchial arch and the eye^{14,15} but are not detectable in the frontonasal region. We therefore hypothesized that Bmp signalling is required for development of maxillary derivatives but not for the frontonasal region. In order to test this hypothesis, we implanted a bead soaked in the Bmp antagonist, Noggin, into the right first branchial arch of a stage-15 chick embryo, adjacent to the presumptive maxillary field (1 mg ml⁻¹, Fig. 1a). Ten days later, an extra protuberance was occasionally seen on the treated side of the upper beak (Fig. 2a). Internally, these abnormal beaks contained a cartilage sheet instead of maxillary bones (Table 1, compare Fig. 2b to 2c, d). Cartilages are normal derivatives of several facial prominences, so it was not clear whether a transformation of maxillary fate had taken place. Endogenous RA

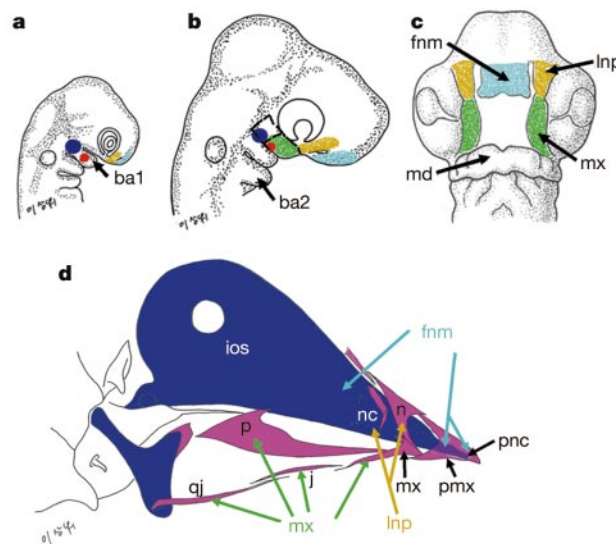


Figure 1 Development of facial prominences and their derivatives. **a**, Stage-15 side view showing position of beads in relation to other structures. The frontonasal mass normally forms medial to the nasal pits and is not seen in this view. Beads are positioned within the first branchial arch (blue bead, noggin bead; red bead, RA bead). Maxillary prominence has not formed yet. **b**, Stage-24 facial prominences side view, green indicates maxilla, yellow indicates the lateral nasal prominence, blue indicates frontonasal mass. Dashed lines indicate proximal and distal regions of maxillary prominence grafted to the limb bud.

Beads end up in the positions indicated. **c**, Stage-24 frontal view. **d**, Skeleton of upper beak, blue indicates cartilage, pink indicates bone. Derivatives of the frontonasal mass, lateral nasal and maxillary prominences are indicated by coloured arrows. ba1, first branchial arch; ba2, second branchial arch; fmm, frontonasal mass; ios, interorbital septum; j, jugal; lnp, lateral nasal prominence; qj, quadratojugal; md, mandibular prominence; mx, maxillary prominence or maxillary bone; n, nasal bone; nc, nasal chonchae; p, palatine; pmx, premaxilla; pnc, prenasal cartilage.

Type of treatment	Total <i>N</i>	Jugal	Maxilla	Palatine	Quadrat	Pterygoid
Noggin, 1 mg ml ⁻¹	21	5	2	4	1	7
Noggin, high concentration*	14	2	1	2	0	9
Noggin + RA†	11	10	9	6	5	8
Noggin + RA, 0.01 mg mg ⁻¹	5	2	1	0	0	0
Noggin + RA 1 mg ml ⁻¹ , beads left in	5	5	5	4	4	4

+ 1 mg ml⁻¹ Noggin + RA 0.1 mg ml⁻¹, both beads left in; or 1 mg ml⁻¹ Noggin + RA 1.0 mg ml⁻¹, RA bead removed at 6 h. These conditions were equivalent.

[illegible]

To test whether RA acts merely by enhancing the effects of Noggin or whether it has independent roles in specifying the identity of the maxillary region, we applied a single bead soaked in a higher concentration of Noggin protein. Application of Noggin-enriched

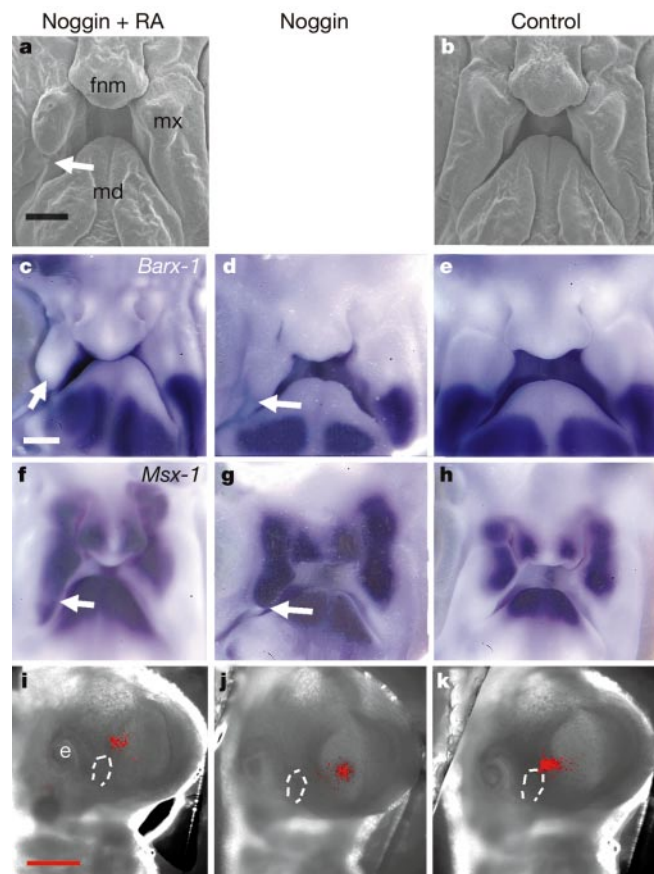


Figure 3 Ontogeny of the duplicated beak phenotype. **a**, SEM front view of stage-29 embryo treated with 1 mg ml⁻¹ Noggin + 1 mg ml⁻¹ RA (removed at 6 h) showing deletion of posterior maxillary prominence (arrow). We notice the comparable size of the frontonasal mass in **a** and in dimethyl sulphoxide (DMSO)-treated control in **b**. **c–e**, *Barx1* expression has been decreased in posterior right maxillary prominence compared to contralateral side (arrows) and compared to controls (**e**). **f–h**, *Msx1* expression has been expanded in posterior right maxillary prominence (arrows) compared to contralateral side and compared to controls (**h**). **i–k**, 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiI) was injected at the same time as the beads were placed and the embryo was fixed 24 h later. The dye has not spread out of the frontonasal mass in treated (**i**, **j**) or control embryos (**k**). Dashed line indicates outline of nasal pit. Scale bar for **a–h** is 500 μm; scale bar for **i–k** is 250 μm. e, eye; fnm, frontonasal mass; md, mandibular prominence; mx, maxillary prominence.

beads alone induced more frequent cartilage formation in the maxillary region than with lower-concentration Noggin beads; however, complete duplications of all the frontonasal mass derivatives were rare (Table 1). Thus RA is required in addition to Noggin to produce ectopic beaks.

The morphology of the beak duplication combined with the loss of maxillary skeletal derivatives suggested that the maxillary prominence was being transformed into a second frontonasal mass. We assessed the morphology of the face 2–4 days after treatment by scanning electron microscopy (SEM, $n = 9$) and in whole mounts ($n = 49$). There were no obvious changes in the size or shape of the frontonasal mass compared to controls (Fig. 3a–h; Noggin + RA, 30/30; Noggin, 15/15; controls, 13/13). This rules out the possibility that the frontonasal mass had split into two fields. In contrast, the morphology of the maxillary prominence was subtly changed and after 4 days it was clear that the posterior half of the maxillary prominence was reduced (Fig. 3a, c, d, f, g). This was confirmed by the failure to upregulate expression of *Barx1* (Fig. 3c, d), which is normally expressed in the posterior half of the maxillary prominence^{18,19} (Fig. 3e, Noggin + RA, 4/5; Noggin, 3/4). *Msx1*, which is expressed in the anterior half of the maxillary prominence²⁰ (Fig. 3h), was induced on schedule in the treated maxillary prominences and ultimately encompassed the whole prominence (Fig. 3f, g; Noggin + RA, 5/5; Noggin, 4/6). This reduction of the posterior maxillary prominence appears to be due to a decrease in cell proliferation rather than an increase in cell death. Cell proliferation was significantly decreased in the posterior than in the anterior maxillary mesenchyme on the treated side (Noggin + RA, $n = 3$, 45%, $P = 0.03$; Noggin, $n = 4$, 40% decrease, $P = 0.01$), and there were no detectable changes in cell death in the posterior maxillary prominences of Noggin + RA-treated embryos (3/3).

We then performed transplantation experiments in order to show whether the treated maxillary prominence, when grown in isolation from the face, could give rise to frontonasal mass derivatives. Forty-eight hours after Noggin + RA treatment, treated and non-treated maxillary prominences were grafted to the wing buds of host embryos. Using tissue from normal embryos, as in previous studies¹¹, we found that none of the grafts of the contralateral, non-treated maxillary prominences formed egg teeth (0/9, Fig. 2e, f) and only one graft contained a very small cartilage nodule. In contrast, egg teeth and cartilage developed in all grafts of treated maxillary prominences (5/5, Fig. 2k, l). Grafts of the mesenchyme proximal to the maxillary prominence (Fig. 2b) gave rise to cartilage sheets similar to the duplicated interorbital septum (2/4). Thus, the duplicated prenasal cartilage and egg tooth came from the maxillary prominence and the interorbital septum came from the mesenchyme proximal to the prominence. These grafting data are consistent with the idea that maxillary mesenchyme had undergone a distinct change in fate.

Finally, to rule out the possibility that frontonasal mass cells had moved into the maxillary process, we labelled small groups of cells in the frontonasal mass with fluorescent dye and studied their expansion patterns. Frontonasal mass cells do not expand into the maxillary region (Noggin + RA, 4/4; Noggin, 5/5; Fig. 3i–k). Thus, the transformation of maxillary prominence identity remains the most plausible explanation for the ectopic beaks.

The signals that mediate the transformation of the maxillary prominence could include Sonic Hedgehog (Shh)²¹. We found that *Shh* was not induced in the mesenchyme by Noggin + RA treatments (data not shown) nor did Shh-soaked beads synergize with Noggin (1/16 from a partial duplication, data not shown). Although there is a role for Shh in later patterning of the frontonasal mass²¹, the specification of frontonasal mass identity is a Shh-independent process.

We looked to see whether *Bmp* expression in the treated maxillary prominence becomes similar to that of the frontonasal mass. *Bmp7* is expressed in frontonasal mass mesenchyme but not in maxillary

mesenchyme at stage 20 (Fig. 4b). In Noggin + RA-treated embryos at stage 20 (24 h after bead implantation) and at stage 24 (48 h after bead implantation) there was a clear increase in expression of *Bmp7* in the maxillary prominence and mesenchyme surrounding the bead (stage 20, 5/5, Fig. 4a; stage 24, 11/12, compare Fig. 4e to 4g). High-concentration Noggin beads also induced a moderate increase in *Bmp7* expression in stage-24 treated maxillas (6/8, Fig. 4i). Thus the maxillary prominence becomes similar to the frontonasal mass in terms of *Bmp7* expression levels and Noggin and RA synergized to induce higher levels of *Bmp7* expression than just Noggin alone. There were also increases in *Bmp2* expression, particularly in the maxillary epithelium (Fig. 4c, h, j). At stage 24, *Bmp2* is normally expressed in the medial maxillary epithelium (Fig. 4d) and at the edges of the frontonasal mass¹⁴ (Fig. 4f, 3/8 Noggin + RA, 3/6 Noggin-treated specimens). There were no detectable changes in *Bmp4* expression (Noggin + RA, 6/6; Noggin, 4/4; data not shown). The ectopic expression of *Bmp2* and *Bmp7* persisted well beyond the time that proteins or RA were being released from the beads^{22,23}. Therefore the data suggest that increased *Bmp7* and *Bmp2* activity at later stages of development is important for frontonasal mass patterning.

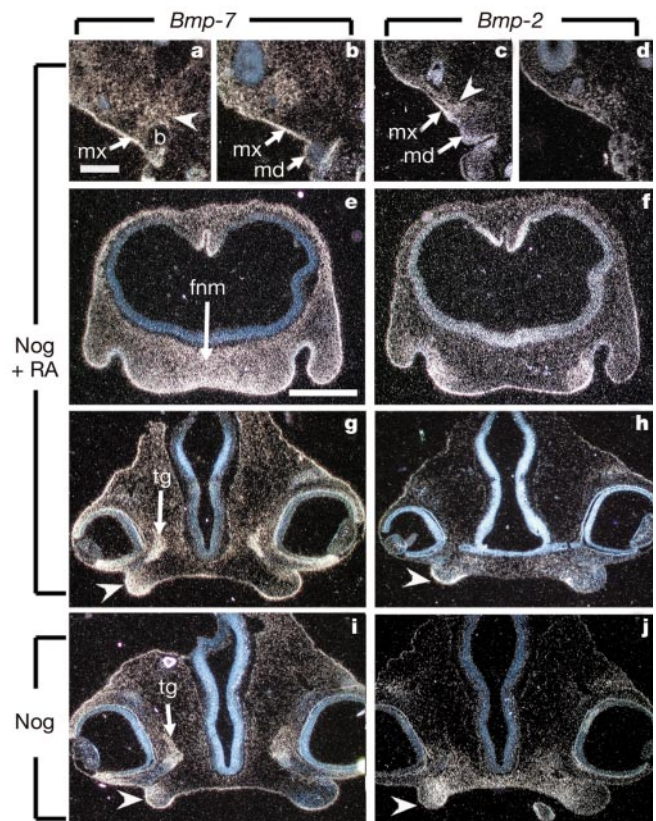


Figure 4 Increased *Bmp7* and *Bmp2* expression in treated maxillary prominences. Embryos in **a–h** were treated on the right side with Noggin 1.0 mg ml^{-1} + 0.1 mg ml^{-1} RA, while embryos in **i** and **j** were treated on the right side with high-concentration Noggin beads. Signal appears as white silver grains on a dark background. Sections in **a** and **b** are right and left sides respectively of a stage-20 embryo (24 h after treatment). Sections in **c** and **d** are adjacent sections of the same embryo. Sections in **e–j** are stage-24 embryos (48 h after bead placement). **a, c**, Increased mesenchymal expression in right maxillary prominence (arrowhead) compared to contralateral side (**b, d**). **e**, Expression of *Bmp7* in the centre and **f**, *Bmp2* at the lateral edges of the frontonasal mass. **g, i**, Increased expression in right maxillary prominence (arrowhead). Normal morphology to trigeminal nerve on treated side compared to the contralateral side. **h**, Increased expression in epithelium (arrowhead). **j**, Increased expression in maxillary mesenchyme (arrowhead). Scale bar for **a–d** is $250 \mu\text{m}$; scale bar for **e–j** is $500 \mu\text{m}$. b, bead; fmm, frontonasal mass; md, mandibular prominence; mx, maxillary prominence; tg, trigeminal ganglion.

Our data suggest a model by which facial prominence identity at stage 15 is specified using different levels of Bmps and retinoids. For example, in order to specify the frontonasal mass, low levels of Bmp and higher levels of RA are required. The small, temporary increase in retinoid level, in combination with the temporary block in Bmp signalling provided by the beads in our study, changed the anterior maxillary prominence into a frontonasal mass rather than a mandible or nasal capsule. At later stages of development, Bmps and retinoids play different roles in patterning within the prominences. Increasing Bmp levels in the maxillary prominence duplicates structures that are normally derived from that region²⁴, whereas blocking retinoid signalling in the frontonasal mass at later stages of development has no effect on morphogenesis¹⁷. We conclude that each prominence acquires its identity separately. Precise control of Bmp and RA levels are involved in specifying at least two of the facial prominences. Furthermore, we have shown that Bmps are used in specifying facial prominence identity just as they are in limb buds¹², digits² and tooth germs¹. The embryonic face is much less accessible and the resulting pattern is more complex than many other parts of the embryo. The ability to initiate frontonasal development in an ectopic location will allow us to determine the genes that encode the identity of the frontonasal mass and to compare the molecular pathways to those used in specifying identity elsewhere in the embryo. □

Methods

Bead implants

Affigel blue beads (150–200 µm, BioRad) were rinsed in phosphate buffered saline (PBS), dried and then soaked in 2 µl of 1 mg ml⁻¹ Noggin protein (Regeneron) at 37 °C for at least 1 h. Higher-concentration beads were prepared by allowing the protein to evaporate and adding another aliquot to the same batch of beads. This was repeated four times for Noggin and five times for Shh (R & D Systems). The Shh beads prepared in this way induced consistent digit duplications when applied to the anterior margin of wing buds. AG1X-2 beads (formate form, BioRad) were soaked in various concentrations of RA dissolved in DMSO as described²⁵. Beads were implanted within the first branchial arch on the right side (Fig. 1a).

Whole-mount *in situ* hybridization and TUNEL staining

The cDNAs were kindly provided by the following individuals: chicken *Msx1* (S. E. Wedden), *Barx1*, *Bmp2*, *Bmp4* (P. Francis-West), *Bmp7* (B. Houston), *Shh* (C. Tabin). Antisense riboprobes were labelled with dig-UTP or ³⁵S-UTP and *in situ* hybridization was performed as described²⁵. TUNEL reaction was carried out on tissue sections 24 h after bead implants as described²⁵ with the exception that a post-fix step was added between the TdT (terminal deoxynucleotidyl transferase) step and the incubation with Dig antibody. In addition, siliconized cover slips were used to prevent tissue tearing.

BrdU staining

Embryos were treated with bead implants at stage 15 and then 24 h later were treated with 10 µM BrdU for 2 h before fixing. Wax sections were treated with 2 M HCl 30 min, 37 °C and 5 µg ml⁻¹ of proteinase K at 37 °C, 10 min, blocked with PTW (PBS + 0.1% Tween-20), 10% goat serum and 0.05% hydrogen peroxide for 30 min, incubated for 1 h at 37 °C with anti-BrdU (Becton-Dickinson, 1:30), followed by biotinylated anti-mouse (1:500, ABC kit, Vectastain) for 1 h and avidin-biotin complex (ABC kit, Vectastain) for 1 h. Detection was performed with 0.05% diaminobenzidine (DAB) and 0.03% hydrogen peroxide in 0.1 M Tris-HCl, pH 7.4. Sections were counterstained with Hoechst 33258 (5 µg ml⁻¹) to visualize the nuclei. Cell counts were made on DAB-stained and Hoechst-stained views to calculate the percentage of proliferating cells in the anterior and posterior maxillary mesenchyme on control and treated sides. A paired *t*-test was performed to determine statistical significance.

Dil injections

Dil was injected into specific locations based on anatomical landmarks at the same time as the beads were implanted. Embryos were fixed after 24 h and examined under rhodamine excitation with a Zeiss axiophot microscope. Bright-field and fluorescent digital images were taken with a Princeton Instruments charge-coupled device (CCD) camera and merged using Adobe Photoshop.

Scanning electron microscopy

SEM was performed on embryos 2–4 days after bead implantation. Embryos were fixed in half-strength Karnovsky's and processed for SEM as described²⁶.

Grafting maxillary prominences to limb buds

Treated and non-treated maxillary prominences from the same embryo were each grafted

to host wing buds¹¹, grown for 7–10 days, fixed in 5% trichloroacetic acid, processed into acetone for 1 day, stained with alcian blue for 2–3 days and cleared in KOH/glycerol.

Skeletal staining of the skull

Skin and eyes were removed from stage 38 embryos; the embryos were fixed in 100% ethanol for 4 days, immersed in 100% acetone for 4 days, rinsed in water and then stained for up to 10 days in alizarin red/alcian blue solution (1 vol. 0.3% alcian blue 8GX in 70% ethanol, 1 vol. 0.1% alizarin red S in 95% ethanol, 1 vol. acetic acid, 17 vol. 70% ethanol). Specimens were cleared in 20% glycerol/2% KOH solution followed by a series of glycerol/H₂O solutions with increasing amounts of glycerol (50%, 80%, 100%).

Received 22 June; accepted 4 October 2001.

1. Tucker, A. S., Matthews, K. L. & Sharpe, P. T. Transformation of tooth type induced by inhibition of BMP signaling. *Science* **282**, 1136–1138 (1998).
2. Drossopoulou, G. *et al.* A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* **127**, 1337–1348 (2000).
3. Haeckel, E. H. P. A. *The Evolution of Man. A Popular Exposition of the Principal Points of Human Ontogeny and Phylogeny* (Appleton, New York, 1886).
4. Le Lièvre, C. S. Participation of neural crest-derived cells in the genesis of the skull in birds. *J. Embryol. Exp. Morph.* **47**, 17–37 (1978).
5. Zimmermann, L. B., De Jesus-Escobar, J. M. & Harland, R. M. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599–606 (1996).
6. Noden, D. M. The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues. *Dev. Biol.* **96**, 144–165 (1983).
7. Couly, G., Grapin-Botton, A., Coltey, P., Ruhin, B. & Le Douarin, N. M. Determination of the identity of the derivatives of the cephalic neural crest: incompatibility between *Hox* gene expression and lower jaw development. *Development* **125**, 3445–3459 (1998).
8. Lumsden, A., Sprawson, N. & Graham, A. Segmental origin and migration of neural crest cells in the hindbrain region of the chick embryo. *Development* **113**, 1281–1291 (1991).
9. Mina, M. & Kollar, E. J. The induction of odontogenesis in non-dental mesenchyme combined with early murine mandibular arch epithelium. *Arch. Oral Biol.* **32**, 123–127 (1987).
10. Hamburger, V. & Hamilton, H. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
11. Richman, J. M. & Tickle, C. Epithelia are interchangeable between facial primordia of chick embryos and morphogenesis is controlled by the mesenchyme. *Dev. Biol.* **136**, 201–210 (1989).
12. Rodríguez-Esteban, C. *et al.* The T-box genes *Tbx4* and *Tbx5* regulate limb outgrowth and identity. *Nature* **398**, 814–818 (1999).
13. Takeuchi, J. K. *et al.* *Tbx5* and *Tbx4* genes determine the wing/leg identity of limb buds. *Nature* **398**, 810–814 (1999).
14. Francis-West, P. H., Tatla, T. & Brickell, P. Expression patterns of the bone morphogenetic protein genes *Bmp-4* and *Bmp-2* in the developing chick face suggest a role in the outgrowth of the primordia. *Dev. Dyn.* **201**, 168–178 (1994).
15. Wall, N. A. & Hogan, B. L. Expression of bone morphogenetic protein-4 (BMP-4), bone morphogenetic protein-7 (BMP-7), fibroblast growth factor-8 (FGF-8) and sonic hedgehog (SHH) during branchial arch development in the chick. *Mech. Dev.* **53**, 383–392 (1995).
16. Niederreither, K., Subbarayan, V., Dolle, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature Genet.* **21**, 444–448 (1999).
17. Schneider, R. A., Hu, D., Rubenstein, J. L., Maden, M. & Helms, J. A. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development* **128**, 2755–2767 (2001).
18. Barlow, A. J., Bogardi, J. P., Ladher, R. & Francis-West, P. H. Expression of chick *Barx-1* and its differential regulation by FGF-8 and BMP signaling in the maxillary Primordia. *Dev. Dyn.* **214**, 291–302 (1999).
19. Shigetani, Y., Nobusada, Y. & Kuratani, S. Ectodermally derived FGF8 defines the maxillomandibular region in the early chick embryo: epithelial-mesenchymal interactions in the specification of the craniofacial ectomesenchyme. *Dev. Biol.* **228**, 73–85 (2000).
20. Brown, J. M. *et al.* Experimental analysis of the control of expression of the homeobox-gene *Msx-1* in the developing limb and face. *Development* **119**, 41–48 (1993).
21. Hu, D. & Helms, J. A. The role of sonic hedgehog in normal and abnormal craniofacial morphogenesis. *Development* **126**, 4873–4884 (1999).
22. Eichele, G., Tickle, C. & Alberts, B. M. Microcontrolled release of biologically active compounds in chick embryos: Beads of 200-µm diameter for the local release of retinoids. *Anal. Biochem.* **142**, 542–555 (1984).
23. Fallon, J. F. *et al.* FGF-2: Apical ectodermal ridge growth signal for chick limb development. *Science* **264**, 104–107 (1994).
24. Barlow, A. J. & Francis-West, P. H. Ectopic application of recombinant BMP-2 and BMP-4 can change patterning and developing chick facial primordia. *Development* **124**, 391–398 (1997).
25. Shen, H. *et al.* Chicken transcription factor AP-2: Cloning, expression and its role in outgrowth of facial prominences and limb buds. *Dev. Biol.* **188**, 248–266 (1997).
26. Richman, J. M. & Tickle, C. Epithelial-mesenchymal interactions in the outgrowth of limb buds and facial primordia in chick embryos. *Dev. Biol.* **154**, 299–308 (1992).

Acknowledgements

We would like to thank Regeneron for generously supplying Noggin protein, L. Niswander, C. Overall, C. Stern, T. Underhill, G. Hunter and V. Diewert for providing critical comments on the manuscript, C. Overall for stimulating discussions, M. A. Ashique for developing the enriching method of bead soaking and A. Wong for help with SEM. This work was funded by a CIHR grant to J.M.R. and salary support from Gyeongsang National University, Chinju, Korea for S.-H.L.

Correspondence and requests for materials should be addressed to J.M.R. (e-mail: richman@interchange.ubc.ca).

Gene defect in ectodermal dysplasia implicates a death domain adapter in development

Denis J. Headon*, Stephanie A. Emmal†, Betsy M. Ferguson†, Abigail S. Tucker‡, Monica J. Justice§, Paul T. Sharpe||, Jonathan Zonana† & Paul A. Overbeek*§

Departments of * Molecular and Cellular Biology and § Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA
† Department of Molecular and Medical Genetics, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA
‡ MRC Centre for Developmental Neurobiology, 4th floor New Hunt's House, King's College, Guy's Hospital, London Bridge, London SE1 1UL, UK
|| Department of Craniofacial Development, GKT Dental Institute, King's College, Guy's Hospital, London Bridge, London SE1 9RT, UK

Members of the tumour-necrosis factor receptor (TNFR) family that contain an intracellular death domain initiate signalling by recruiting cytoplasmic death domain adapter proteins^{1,2}. Edar is a death domain protein of the TNFR family that is required for the development of hair, teeth and other ectodermal derivatives^{3,4}. Mutations in Edar—or its ligand, Eda—cause hypohidrotic ectodermal dysplasia in humans and mice^{3–7}. This disorder is characterized by sparse hair, a lack of sweat glands and malformation of teeth⁸. Here we report the identification of a death domain adapter encoded by the mouse *crinkled* locus. The *crinkled* mutant has an hypohidrotic ectodermal dysplasia phenotype identical to that of the *edar* (*downless*) and *eda* (*Tabby*) mutants⁹. This adapter, which we have called Edaradd (for Edar-associated death domain), interacts with the death domain of Edar and links the receptor to downstream signalling pathways. We also identify a missense mutation in its human orthologue, *EDARADD*, that is present in a family affected with hypohidrotic ectodermal dysplasia. Our findings show that the death receptor/adapter signalling mechanism is conserved in developmental, as well as apoptotic, signalling.

Members of the TNFR family initiate intracellular signalling by two distinct mechanisms. Most family members lack a death domain and directly recruit members of the TNF-receptor-associated factor (Traf) family to initiate signalling; alternatively, receptors containing a death domain rely on cytoplasmic death domain adapters to mediate the formation of a signalling complex¹⁰.

At present, this death domain receptor group consists of eight members: p75 neurotrophin receptor (p75NTR), TNFR1, Fas, DR3, DR4, DR5, DR6 and Edar². The principal function of most of these proteins is to induce apoptosis. By contrast, Edar regulates cell fate and morphogenesis during ectodermal development^{3,11}. Two death domain adapters, TNF-receptor-associated death domain (Tradd)¹² and Fas-associated death domain (Fadd)¹³, have been found to be crucial to death receptor signalling. Tradd and Fadd interact with one another and with the death domains of all death receptors except Edar and p75NTR, thus serving as general adapters for the death receptors^{12–16}. No death domain adapters that interact with p75NTR or Edar have been identified.

Tabby (*Ta*), *downless* (*dl*) and *crinkled* (*cr*) mutant mice display an identical ectodermal dysplasia phenotype caused by failure of hair follicle induction and defective morphogenesis of teeth and glands¹⁷. The *Ta* gene encodes Eda, which is a ligand for Edar^{11,18}—the product of the *dl* gene. To identify components of the Edar signalling pathway, we initiated the positional cloning of the *cr* gene, which maps close to the *nidogen* locus on chromosome 13 (ref. 19). We used microsatellite markers from this region to genotype roughly 700 meioses from a *cr* × *Mus castaneus* backcross, and then generated a yeast artificial chromosome (YAC) contig across a region defined by two markers that did not recombine with *cr* (Supplementary Information). We identified genes from this region by using a combination of complementary DNA selection and comparison of the selected cDNAs with the draft human genome sequence. The amino-acid sequence of each candidate gene was searched for domains that resemble those involved in TNF signal transduction. This approach identified a death domain encoded in an unfinished bacterial artificial chromosome (BAC) sequence from human chromosome 1q42.2–43, a region for which conservation of synteny with the *cr* region of mouse chromosome 13 has been established²⁰.

Oligonucleotides from this human death domain sequence were used to amplify the corresponding mouse transcript from fetal skin RNA. The mouse cDNA contains a 627-nucleotide open reading frame (ORF). The predicted 208-amino-acid protein includes a carboxy-terminal death domain that is most similar to the death domain of MyD88 (Fig. 1a), a cytoplasmic transducer of Toll/interleukin receptor signalling²¹. The amino-terminal region contains the sequence Pro-Ile-Gln-Asp-Thr, which corresponds to the Traf interaction consensus sequence Pro-X-Gln-X-Thr (ref. 22). This cDNA hybridized to a 6-kilobase (kb) messenger RNA in wild-type but not *cr* fetal skin by northern blotting (Fig. 1b). Polymerase chain reaction (PCR) analysis of genomic DNA showed that the entire coding region is deleted from the

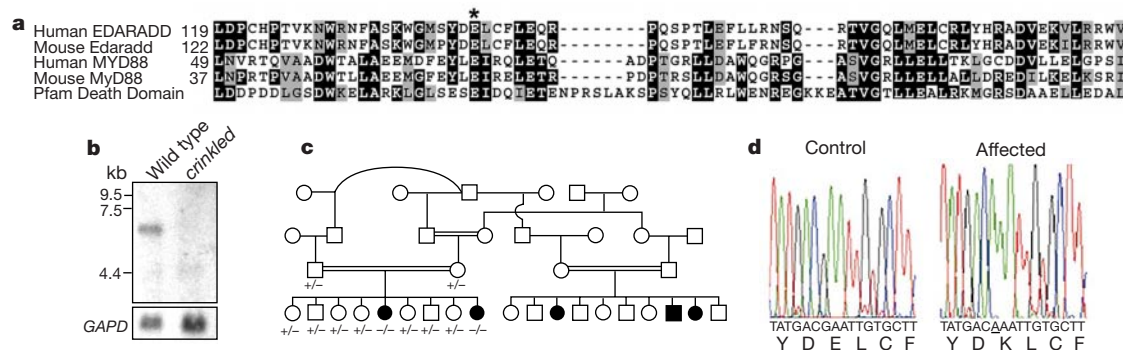


Figure 1 Edaradd death domain alignment and mutation analyses. **a**, Alignment of EDARADD and MYD88 death domains. Asterisk indicates the glutamate residue mutated in human family ED1176. The Pfam death domain consensus is also shown. Identical amino acids are shaded in black, similar ones in grey. **b**, Northern blot of E18.5 mouse skin RNA probed with *edaradd*. Wild type shows a 6-kb transcript, but no transcript is

detected in the *cr* mutant. The amount of mRNA in each lane is indicated by the *GAPD* signal. **c**, Pedigree of HED-affected family ED1176. Individuals with HED are depicted as filled circles or squares. All affected individuals from whom DNA was available were homozygous for the G424→A mutation (–/–). **d**, Automated sequence traces showing the G→A mutation in an individual affected by HED.

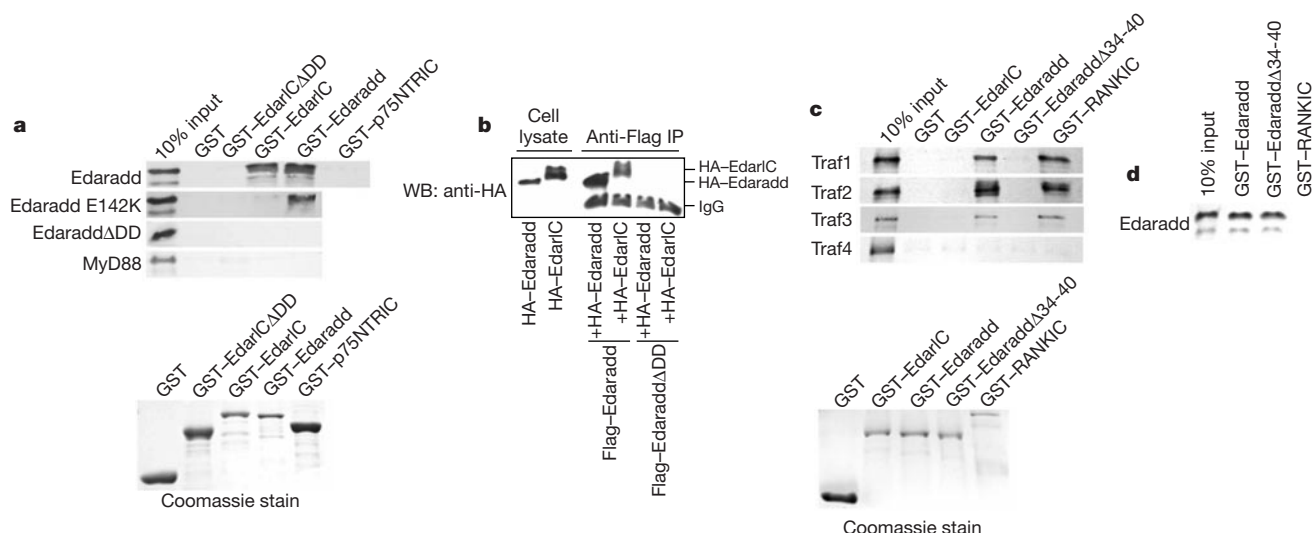


Figure 2 Interaction between Edaradd and Edar. **a**, 35 S-labelled Edaradd interacts with GST fused to the Edar intracellular domain (GST-EdarC) and with Edaradd itself (GST-Edaradd), but not with GST alone, an Edar death domain deletion mutant (GST-EdarCΔDD), or with the p75NTR intracellular domain (GST-p75NTRIC). The Edaradd Glu142→Lys mutant (EdaraddE142K) has reduced interaction with wild-type Edaradd and greatly diminished interaction with EdarC. A truncated Edaradd lacking the death domain (EdaraddΔDD) and MyD88 fail to interact with any of the GST fusions. **b**, HA-tagged EdarC and Edaradd co-precipitate with Flag-tagged Edaradd. Removing the

Edaradd death domain abolishes association with EdarC and Edaradd. **c**, GST pull-down assays for Traf interaction with EdarC, Edaradd, EdaraddΔ34-40 and RANKIC. Traf1, -2 and -3 interact with Edaradd and RANKIC. Traf4 does not interact with any of these proteins. Deleting the Edaradd interaction consensus (EdaraddΔ34-40) abolishes interaction with the Trafs. Coomassie staining shows the amount of each GST fusion protein used. **d**, 35 S-labelled Edaradd interacts with GST-EdaraddΔ34-40 and GST-Edaradd, but not with GST-RANKIC. In **a**, **c**, **d**, the GST fusion proteins are shown on top, and the 35 S-labelled proteins that were tested for interaction are indicated on the left.

genome of the *cr* mutant (Supplementary Information).

The predicted human protein is 80% identical to the murine protein, with almost complete identity in the death domain and a completely conserved Traf-binding consensus site (Fig. 1a; and Supplementary Information). We sequenced genomic DNA from human families affected with autosomal hypohidrotic ectodermal dysplasia (HED) to look for variations in the *EDARADD* gene. A large consanguineous family (ref 23; and Fig. 1c) was found to have the mutation G424→A (Fig. 1d), which changes a glutamate residue to a lysine (Glu142→Lys) in the EDARADD death domain. This alters the charge of an amino acid that is conserved between mouse and human EDARADD, as well as between the MyD88 and the Pfam death domain consensus sequence (Fig. 1a). This mutation cosegregated with the HED phenotype, with all affected individuals displaying homozygosity for the mutation. The 424G→A variant was not found in control DNA samples from 100 individuals (200 chromosomes) of diverse racial and ethnic backgrounds.

The mutant phenotypes and domain structures of Edar and Edaradd suggest that they physically interact. By using a glutathione S-transferase (GST) pull-down assay, we found that Edaradd interacts with the intracellular domain of Edar (EdarC) (Fig. 2a), but not with an Edar mutant in which the intracellular domain is truncated before the death domain (EdarCΔDD), or with the p75NTR intracellular domain (p75NTRIC) (Fig. 2a). The EdarCΔDD truncation corresponds to the dominant-negative *Df^{sleek}* allele in mouse³. Edaradd also self-associates (Fig. 2a), a property common to many death domain proteins¹⁰ including Tradd¹².

As the human and mouse Edaradd proteins display 96% identity in their death domains, we engineered the Glu142→Lys mutation into mouse Edaradd to make a protein analogous to the human variant found in the HED-affected family (Fig. 1d). This mutation moderately diminished Edaradd self-association and greatly reduced its ability to bind to Edar (Fig. 2a). Deletion of the Edaradd death domain resulted in an inability to interact with Edar or Edaradd (Fig. 2a). MyD88 did not interact with Edar or Edaradd

(Fig. 2a). We tested for Edar and Edaradd interaction in a cellular context by immunoprecipitating epitope-tagged proteins expressed in cultured 293T mammalian cells. For this assay we used the Edar intracellular domain, rather than the full-length receptor, as we could not detect the epitope-tagged full-length form by western blotting. We found that EdarC co-precipitated with Edaradd, and that deleting the Edaradd death domain abolished this interaction (Fig. 2b).

The consensus sequence Pro-X-Gln-X-Thr is used by several non-death domain TNFRs as a docking site for Traf1, -2, -3 and -5 (ref. 22). As Edaradd contains this consensus site, we performed GST pull-down assays to determine whether it can directly interact with the Trafs. We used the intracellular domain of RANK, a TNFR family member that lacks a death domain and directly interacts with Traf1, -2, -3, -5 and -6 (ref. 24), as a positive control. Edaradd was found to associate with Traf1, -2 and -3, but not with Traf4 (Fig. 2c). Deleting seven amino acids corresponding to the Edaradd Traf interaction consensus site (EdaraddΔ34-40) abolished Traf binding (Fig. 2c), but EdaraddΔ34-40 retained the ability to interact with wild-type Edaradd (Fig. 2d).

We compared the expression patterns of *edar* and *edaradd* in fetal mouse tissues affected by their mutation. *In situ* hybridization to serial sections showed that these genes are co-expressed in epithelial cells during the formation of hair follicles (Fig. 3a–f) and teeth (Fig. 3g–i), supporting the idea that Edar and Edaradd interact in a common signalling pathway. The *Shh* and *edar* genes are expressed in hair follicle progenitor cells before and during follicle morphogenesis^{3,25}, and both genes require Edar signalling for focal expression³. Thus, at embryo day (E) 15.5, the skin of a wild-type mouse is dotted with clusters of *Shh*-expressing (Fig. 3j) and *edar*-expressing (Fig. 3k) cells. In *cr* mutants, however, *Shh* (Fig. 3l) and *edar* (Fig. 3m) are detected only in the whisker follicles around the mouth, which develop normally in the mutant mice.

Death receptors can activate NF-κB as part of a cell survival pathway, and Edar also activates NF-κB (refs 18, 26, 27). We found that endogenous Edaradd is expressed in the HEK293T cell line

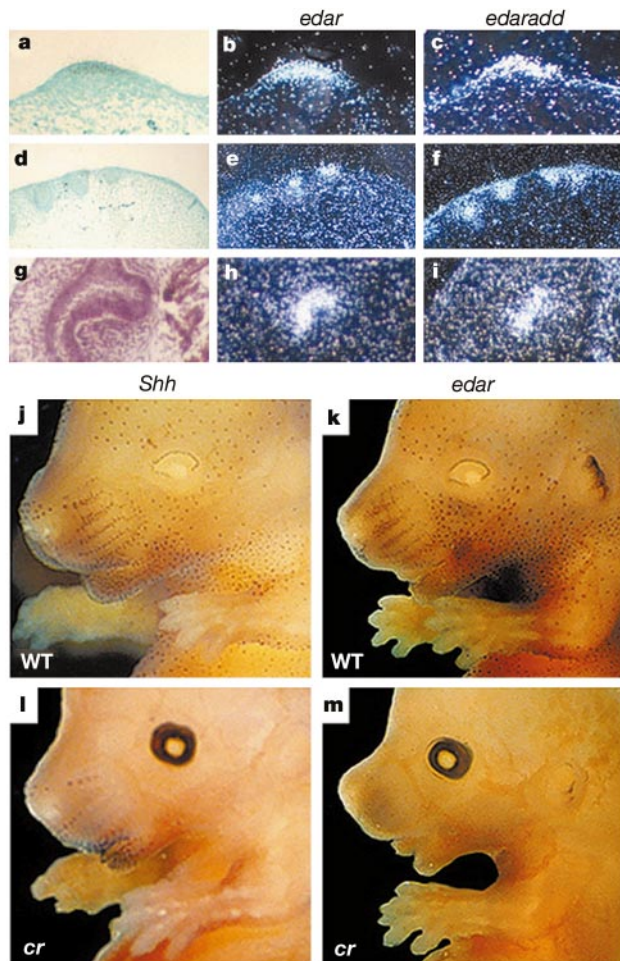


Figure 3 Co-expression of *edar* and *edaradd*. **a–i**, Serial sections of E12.5 skin (**a–c**), E14.5 skin (**d–f**) and E15.5 tooth (**g–i**) either stained with malachite green (**a, d**) or haematoxylin (**g**), or hybridized to *edar* (**b, e, h**) or *edaradd* (**c, f, i**) riboprobes. *edar* and *edaradd* are co-expressed in epithelial cells in developing whisker hair follicles (**a–f**) and in the enamel knot of the developing tooth (**g–i**). **j–m**, Whole-mount *in situ* hybridization to E15.5 embryos. *Shh* (**j, l**) and *edar* (**k, m**) are expressed in whisker follicles on the snout and in hair follicle precursors on the head and trunk of wild-type embryos (**j, k**). No hair follicle induction is detected at this age in the *cr* mutant (**l, m**). The *cr* mutant has black eyes as it is on a pigmented genetic background.

(Fig. 4a), and that its overexpression activates an NF- κ B reporter gene in a dose-dependent manner (Fig. 4b, c). The Glu142→Lys mutation reduces the activation of NF- κ B by Edaradd (Fig. 4b); however, removing the Traf-binding motif has little effect on NF- κ B activation (Fig. 4b), suggesting that Traf recruitment is not crucial for this function. Edaradd signalling was abolished by deletion of its N-terminal 70 amino acids (Edaradd Δ N70) (Fig. 4d). When co-expressed with full-length Edar, Edaradd Δ N70 interferes with the receptor's ability to activate NF- κ B (Fig. 4d), presumably acting as a dominant-negative inhibitor. That Edaradd Δ N70 can block the activation of NF- κ B by Edar illustrates Edaradd's central role in Edar signalling, and also suggests that it has a modular structure¹⁰ in which the C-terminal death domain is required for receptor engagement and the N-terminal region is responsible for signal transduction.

Our results identify a pathway in which Edar is activated by Eda and uses Edaradd as an adapter to build an intracellular signal-transducing complex. This linear pathway explains the identical phenotypes of the *Ta*, *dl* and *cr* mutants, and also the genetic heterogeneity of human hypohidrotic ectodermal dysplasia. The adapter molecules Tradd and Fadd are central to death receptor

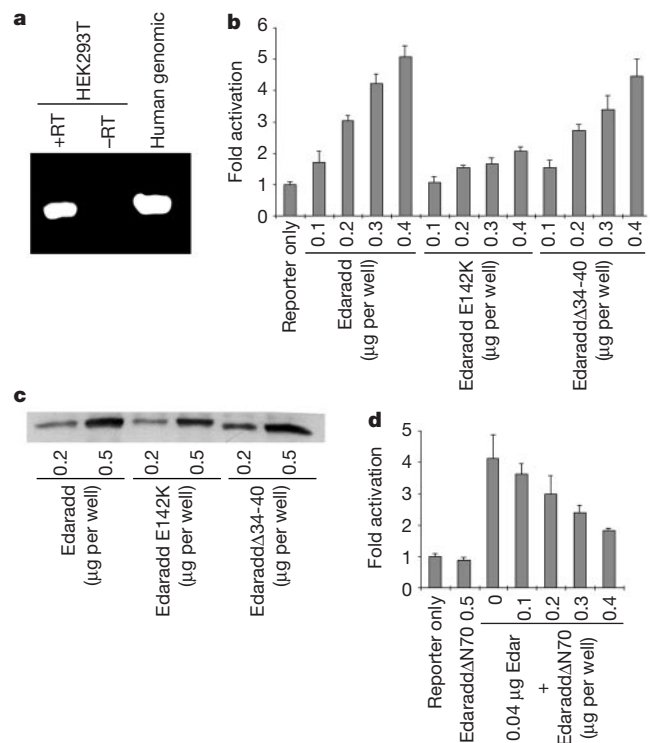


Figure 4 Edaradd mediates Edar activation of NF- κ B. **a**, PCR with reverse transcription (RT) showing *EDARADD* expression in untransfected HEK293T cells. **b**, Edaradd overexpression induces an NF- κ B luciferase reporter gene in a dose-dependent manner. The Glu142→Lys (E142K) mutation exhibits weak stimulation of NF- κ B, but Edaradd Δ 34-40 shows little change in NF- κ B activation. The amount of Edaradd expression vector transfected per well is indicated. **c**, Western blot showing relative expression levels of transfected HA-tagged Edaradd, Edaradd Gly142→Lys, and Edaradd Δ 34-40. **d**, Overexpression of full-length Edar activates NF- κ B, but co-expression of Edaradd Δ N70 blocks this activation. For each co-transfection, 0.04 μ g of Edar expression vector was transfected per well. The amount of Edaradd Δ N70 expression vector used is indicated. Edaradd Δ N70 alone does not activate NF- κ B (second bar).

signalling^{1,2}, and the molecular identity of Edaradd illustrates the conservation of the archetypal death receptor pathway in a developmental context. □

Methods

edaradd mapping and identification

We obtained *cr* mice on a C57/C3H background from The Jackson Lab. *CastEi* mice were a gift from A. Bradley. Microsatellite specific oligonucleotides *D13Mit44*, *D13Mit57*, *D13Mit114* and *D13Mit216*, and YACs were obtained from Research Genetics. We carried out cDNA selection as described³.

The sequences of cDNA clones were searched against the high throughput genomic sequence (htgs) database using BLAST at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/blast>). Matching human BAC sequences were input into the *FGENESH* gene prediction program at the Sanger Centre website (<http://genomic.sanger.ac.uk/gf/gf.shtml>). The amino-acid sequences of predicted genes were searched for protein domains using PFAM v3.2 at the San Diego Supercomputer Center website (<http://fps.sdsc.edu>).

We carried out rapid amplification of cDNA ends using the Generacer procedure (Invitrogen) with E17.5 mouse skin cDNA as the template. Oligonucleotide sequences and PCR conditions are available on request.

Mutation analysis

Families with autosomal recessive hypohidrotic ectodermal dysplasia were recruited into a research study approved by the institutional review board of the Oregon Health Sciences University. Consent was obtained for the use of clinical information, relevant family history and DNA samples. We isolated genomic DNA from patients by described methods²⁸. PCR fragments from each of the six exons of *EDARADD* were amplified and sequenced. Genomic DNA from relevant family members, as well as from control individuals, was analysed for the presence or absence of variants by ASO analysis²⁸ or by

direct sequencing. Oligonucleotide sequences and PCR conditions are available on request.

Northern and *in situ* hybridizations

Northern and *in situ* hybridization to sectioned tissue were done as described³¹. The 738-nucleotide *edaradd* probe contained the entire 627-nucleotide ORF plus 76 nucleotides of the 5' and 35 nucleotides of the 3' untranslated region. For whole-mount *in situ* hybridization³, we used full-length *edar* and *Shh* riboprobes.

Protein interaction and NF- κ B reporter assays

GST fusion constructs were prepared by cloning cDNAs into pGEX4T-1 (Amersham Pharmacia). The intracellular portion of Edar (EdarIC) contained amino acids 212–448; the Edar death domain deletion (EdarICADD) contained amino acids 212–322; Edar-addADD contained amino acids 1–124. Deletion of the seven amino acids Tyr-Pro-Ile-Gln-Asp-Thr-Gly from wild-type Edaradd created Edaradd Δ 34–40. RANKIC is amino acids 235–625 and p75NTRIC amino acids 264–417 of mouse RANK and mouse p75NTR, respectively. We carried out *in vitro* transcription/translation using the TNT Coupled Reticulocyte Lysate system (Promega) in the presence of [³⁵S]methionine (Amersham Pharmacia). We then expressed the GST fusion proteins in *Escherichia coli* strain BL21. Bacteria were lysed by sonication in PBS plus 0.5% Triton X-100. We used 2 μ l of *in vitro* transcription/translation reaction for each pull-down assay. Protein mixtures were incubated at 4 °C for 2 h with 15 μ l of glutathione–sepharose beads (Amersham Pharmacia). The beads were washed four times with PBS plus 0.1% Triton X-100 and bound proteins were eluted with 20 μ l of 50 mM glutathione.

For immunoprecipitations, individual cDNAs were cloned into pSG5 (Stratagene) downstream of two copies of a Flag or haemagglutinin A (HA) epitope tag. HEK293T cells grown in 25-cm² flasks were transfected using Lipofectamine2000 (Lifetech). Whole-cell extracts were prepared 30 h after transfection by adding 1 ml of lysis buffer (50 mM HEPES, 150 mM NaCl, 0.5 % NP-40, protease inhibitors). Cell lysates (850 μ l) were added to 40 μ l of anti-Flag M2 affinity gel (Sigma) and rotated at 4 °C overnight. The affinity gels were washed with lysis buffer five times and eluted with 20 μ l of 0.1 M glycine (pH 3.1); the eluted proteins were assayed on western blots using a 1:1,250 dilution of rat anti-HA primary antibody (3F10; Roche) followed by a 1:3,500 dilution of HRP conjugated rabbit anti-rat secondary antibody (Zymed) and detection by enhanced chemiluminescence (Amersham).

For NF- κ B activation assays, HEK293T cells grown in 24-well plates were transfected using Lipofectamine2000 (Lifetech). Each well received 0.2 μ g of pNF- κ B luciferase reporter (Clontech), cDNA(s) for expression in pSG5–HA, and empty pSG5–HA vector to give 1.0 μ g of DNA per well. We lysed the cells 28 h after transfection and measured luciferase activities (Tropix; Promega) using a luminometer. Each transfection was performed in quadruplicate. To compare relative protein expression levels, transfections were done as in the reporter assays. After 28 h, each well of cells was lysed in 100 μ l of SDS–PAGE buffer, and 20 μ l of lysate was used for anti-HA western blotting as described above.

Received 31 May; accepted 11 October 2001.

1. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* **281**, 1305–1308 (1998).
2. Locksley, R. M., Killeen, N. & Lenardo, M. J. The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell* **104**, 487–501 (2001).
3. Headon, D. J. & Overbeek, P. A. Involvement of a novel Tnf receptor homologue in hair follicle induction. *Nature Genet.* **22**, 370–374 (1999).
4. Monreal, A. W. *et al.* Mutations in the human homologue of mouse *dl* cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nature Genet.* **22**, 366–369 (1999).
5. Kere, J. *et al.* X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nature Genet.* **13**, 409–416 (1996).
6. Ferguson, B. M. *et al.* Cloning of *Tabby*, the murine homologue of the human *EDA* gene: evidence for a membrane associated protein with a short collagenous domain. *Hum. Mol. Genet.* **6**, 1589–1594 (1997).
7. Srivastava, A. K. *et al.* The *Tabby* phenotype is caused by mutation in a mouse homologue of the *EDA* gene that reveals novel mouse and human exons and encodes a protein (ectodysplasin-A) with collagenous domains. *Proc. Natl Acad. Sci. USA* **94**, 13069–13074 (1997).
8. Clarke, A. Hypohidrotic ectodermal dysplasia. *J. Med. Genet.* **24**, 659–663 (1987).
9. Falconer, D. S., Fraser, A. S. & King, J. W. B. The genetics and development of “crinkled”, a new mutant in the mouse. *J. Genet.* **50**, 324–344 (1951).
10. Hofmann, K. The modular nature of apoptotic signaling proteins. *Cell. Mol. Life Sci.* **55**, 1113–1128 (1999).
11. Tucker, A. S. *et al.* Edar/Eda interactions regulate enamel knot formation in tooth morphogenesis. *Development* **127**, 4691–4700 (2000).
12. Hsu, H., Xiong, J. & Goeddel, D. V. The TNF receptor 1-associated protein TRADD signals cell death and NF- κ B activation. *Cell* **81**, 495–504 (1995).
13. Chinnaiyan, A. M., O'Rourke, K., Tewari, M. & Dixit, V. M. FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell* **81**, 505–512 (1995).
14. Chaudhary, P. M. *et al.* Death receptor 5, a new member of the TNFR family, and DR4 induce FADD-dependent apoptosis and activate the NF- κ B pathway. *Immunity* **7**, 821–830 (1997).
15. Chinnaiyan, A. M. *et al.* Signal transduction by DR3, a death domain-containing receptor related to TNFR-1 and CD95. *Science* **274**, 990–992 (1996).
16. Pan, G. *et al.* Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. *FEBS Lett.* **431**, 351–356 (1998).
17. Lyon, M. F., Rastan, S. & Brown, S. D. M. *Genetic Variants and Strains of the Laboratory Mouse* (Oxford University Press, Oxford, 1996).
18. Yan, M. *et al.* Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**, 523–527 (2000).

19. Blake, J. A. *et al.* The mouse genome database (MGD): Integration nexus for the laboratory mouse. *Nucleic Acids Res.* **29**, 91–94 (2001).
20. Perou, C. M. *et al.* Comparative mapping in the beige-satin region of mouse chromosome 13. *Genomics* **39**, 136–146 (1997).
21. Medzhitov, R. *et al.* MyD88 is an adaptor protein in the Toll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
22. Inoue, J. *et al.* Tumor necrosis factor receptor-associated factor (TRAF) family: Adapter proteins that mediate cytokine signaling. *Exp. Cell Res.* **254**, 14–24 (2000).
23. Munoz, F. *et al.* Definitive evidence for an autosomal recessive form of hypohidrotic ectodermal dysplasia clinically indistinguishable from the more common X-linked disorder. *Am. J. Hum. Genet.* **61**, 94–100 (1997).
24. Galibert, L., Tometsko, M. E., Anderson, D. M., Cosman, D. & Dougall, W. C. The involvement of multiple tumor necrosis factor receptor (TNFR)-associated factors in the signaling mechanisms of receptor activator of NF- κ B, a member of the TNFR superfamily. *J. Biol. Chem.* **273**, 34120–34127 (1998).
25. Bitgood, M. J. & McMahon, A. P. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126–138 (1995).
26. Kumar, A., Eby, M. T., Sinha, S., Jasmin, A. & Chaudhary, P. M. The ectodermal dysplasia receptor activates the nuclear factor- κ B, JNK, and cell death pathways and binds to ectodysplasin A. *J. Biol. Chem.* **276**, 2668–2677 (2000).
27. Döflinger, R. *et al.* X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF- κ B signaling. *Nature Genet.* **27**, 277–285 (2001).
28. Ferguson, B. M. *et al.* Scarcity of mutations detected in families with X linked hypohidrotic ectodermal dysplasia: diagnostic implications. *J. Med. Genet.* **35**, 112–115 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank the families and physicians who assisted in gathering clinical information and samples, in particular G. Lestringant. We also thank D. Moore and D. Dowhan for technical advice and reagents. This work was supported by grants from the NIH and the National Foundation for Ectodermal Dysplasias.

Correspondence and requests for materials should be addressed to P.A.O. (e-mail: overbeek@bcm.tmc.edu). The cDNAs have been deposited in GenBank (accession codes AY028914, human *EDARADD*; and AF358671, mouse *edaradd*).

Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage

Akio Ohta & Michail Sitkovsky

Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 10 Center Drive, Room 10/11N311, Bethesda, Maryland 20892-1892, USA

Inappropriate or prolonged inflammation is the main cause of many diseases¹; for this reason it is important to understand the physiological mechanisms that terminate inflammation *in vivo*². Agonists for several G_s-protein-coupled receptors³, including cell-surface adenosine purinergic receptors^{4–7}, can increase levels of immunosuppressive cyclic AMP in immune cells^{8–15}; however, it was unknown whether any of these receptors regulates inflammation *in vivo*. Here we show that A2a adenosine receptors have a non-redundant role in the attenuation of inflammation and tissue damage *in vivo*. Sub-threshold doses of an inflammatory stimulus^{16,17} that caused minimal tissue damage in wild-type mice were sufficient to induce extensive tissue damage, more prolonged and higher levels of pro-inflammatory cytokines, and death of male animals deficient in the A2a adenosine receptor. Similar observations were made in studies of three different models of inflammation and liver damage as well as during bacterial endotoxin-induced septic shock. We suggest that A2a adenosine receptors are a critical part of the physiological

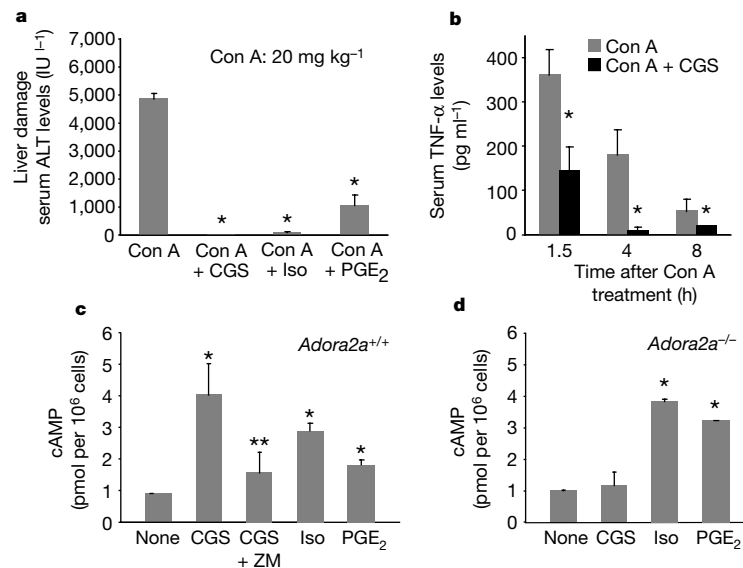


Figure 1 Pharmacological activation of Adora2a or other G_s -protein-coupled receptors *in vivo* prevents Con A-induced liver damage and pro-inflammatory TNF- α accumulation *in vivo*. **a**, Activation of any of several cAMP-elevating G_s -coupled receptors can block Con A-induced liver injury. Female C57BL/6 mice ($n = 5$) were first injected with Adora2a agonist CGS21680 (CGS) (2 mg kg^{-1}), 100 mg kg^{-1} isoproterenol (Iso), or 5 mg kg^{-1} PGE $_2$, and then with Con A (20 mg kg^{-1}). The Con A-induced liver damage was evaluated after 8 h. The differences between treated and untreated mice are statistically significant as indicated by the asterisk ($P < 0.05$). **b**, CGS21680 (2 mg kg^{-1}) inhibits Con A-induced elevation of serum TNF- α . **c**, **d**, Adora2a agonist CGS21680 does not increase cAMP in

mononuclear cells from Adora2a $^{-/-}$ mice livers. Adora2a antagonist ZM241385 inhibits CGS21680-induced cAMP in mononuclear cells from Adora2a $^{+/+}$ mice livers. Lymphoid cells from Adora2a $^{-/-}$ mice retain the cAMP response to ligands of other G_s -protein-coupled receptors. Cells were treated *in vitro* with $10 \mu\text{M}$ CGS21680, $1 \mu\text{M}$ ZM241385 (ZM), $100 \mu\text{M}$ isoproterenol, or $1 \mu\text{M}$ PGE $_2$ for 30 min at 37°C . Data shown represent the mean $\pm$ standard error of three preparations. Asterisk, $P < 0.05$ compared with controls; double asterisk, $P < 0.05$ compared with cells incubated with CGS21680. Units on the y axis of Figs 1a, 2a, c and 3 are International Units of ALT activity.

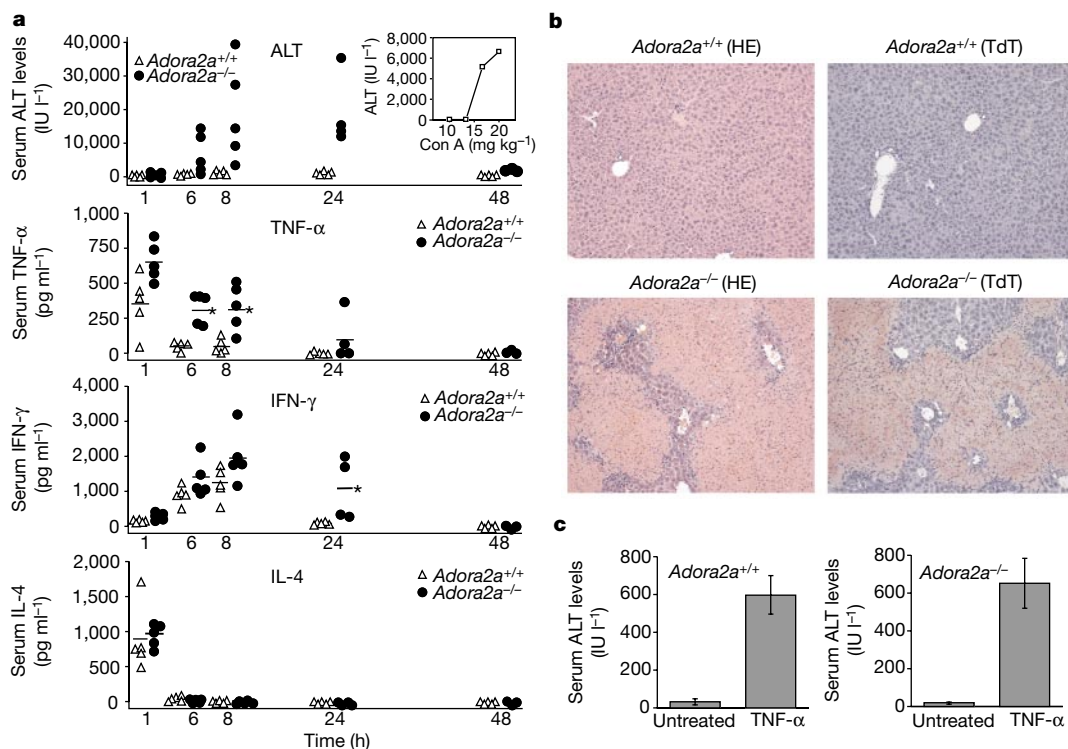


Figure 2 Enhanced and prolonged accumulation of pro-inflammatory cytokines and exaggerated liver damage in A2a-receptor-deficient mice. **a**, Adora2a $^{+/+}$ ($n = 5$) and Adora2a $^{-/-}$ ($n = 5$) mice were injected intravenously with a sub-optimal dose (12.5 mg kg^{-1}) of Con A (inset, ALT graph). Serum levels of cytokines were measured at indicated times. Data shown are representative of four separate experiments. Asterisk, $P < 0.05$ compared with Adora2a $^{+/+}$ mice. **b**, Increased inflammation and liver damage in

Adora2a $^{-/-}$ mice as evidenced by accumulation of dead cells and leukocytes using TdT apoptosis test and haematoxylin and eosin (HE) stain. **c**, A deficiency in A2a receptors does not affect the susceptibility of hepatocytes to *in vivo* damage by TNF- α . Adora2a $^{+/+}$ ($n = 5$) and Adora2a $^{-/-}$ ($n = 5$) mice were treated with a combination of D-galactosamine (700 mg kg^{-1}) and TNF- α ($15 \mu\text{g kg}^{-1}$). Liver damage was evaluated 6 h later by measuring serum ALT levels.

negative feedback mechanism for limitation and termination of both tissue-specific and systemic inflammatory responses.

The accumulation of extracellular adenosine in inflamed and damaged tissues^{18–22} and the immunosuppressive properties of cAMP-elevating adenosine receptors^{8–11,14,15,23} indicate that signalling by A2a adenosine receptors on immune cells is a possible natural mechanism of inhibition and/or termination of inflammation. To test whether the absence of functional A2a adenosine receptors results in increased inflammation and exacerbated tissue damage, we used mice deficient in the adenosine A2a receptor gene (*Adora2a*^{−/−})^{24,25}, which had been characterized previously in studies of immune cells^{26,27} and cardiovascular and neurological systems^{24,25}.

Concanavalin A (Con A)-induced liver injury in mice is mediated by T cells, macrophages and cytokines tumour-necrosis factor (TNF)-α, interleukin (IL)-4 and interferon (IFN)-γ, and represents a well described *in vivo* inflammation model of viral and autoimmune hepatitis^{16,17}. The pharmacological activation of A2a receptors by a selective A2a receptor agonist²⁶, CGS21680, prevented liver damage (Fig. 1a and Supplementary Information) and pro-inflammatory TNF-α accumulation *in vivo* (Fig. 1b). CGS21680 also inhibited secretion of TNF-α, IL-12 and IFN-γ by activated macrophages and T cells *in vitro* (data not shown). These data, however, do not show conclusively that adenosine receptors actually serve *in vivo* as physiological regulators of inflammation; other G_s-coupled receptors (for example, prostaglandin E₂ (PGE₂), histamine H₂ and β-adrenergic receptors^{12,13}) could also function in a similar capacity, as they trigger cAMP accumulation (Fig. 1c, d), inhibit cytokine secretion by immune cells *in vitro* and prevent liver damage *in vivo* (Fig. 1a).

We next investigated whether genetic inactivation of A2a receptors increases the intensity and prolongs duration of T-cell- and macrophage-dependent pro-inflammatory cytokine accumulation and liver damage *in vivo* in *Adora2a*^{−/−} mice treated with Con A. We

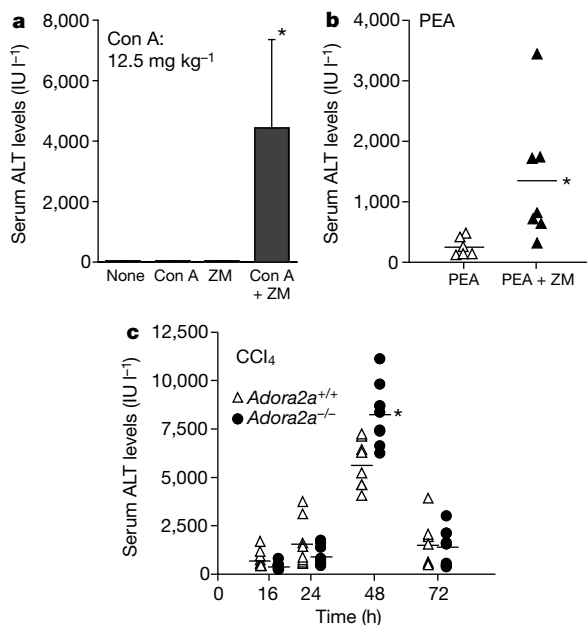


Figure 3 Pharmacological inactivation of A2a receptors *in vivo* by antagonist ZM241385 exacerbates Con A- (a) and PEA-induced (b) liver damage. Genetic *Adora2a* deficiency enhances carbon tetrachloride (CCl₄) hepatotoxicity (c). B6 mice (*n* = 5) were injected with 12.5 mg kg^{−1} Con A alone or in combination with *Adora2a* antagonist ZM241385 (2 mg kg^{−1}) followed by ALT measurements and histological evaluation. Exotoxin A (PEA, 100 μg kg^{−1}) was injected alone (*n* = 6) or in combination with ZM241385 (*n* = 7), and liver damage was evaluated at 24 h. *Adora2a*^{+/+} (*n* = 7) and *Adora2a*^{−/−} mice (*n* = 7) were injected intraperitoneally with 0.5 ml kg^{−1} CCl₄. Data shown are representative of three separate experiments. Asterisk, *P* < 0.05 compared with *Adora2a*^{+/+} mice.

confirmed that A2a receptor signalling (cAMP accumulation) is abolished in liver mononuclear cells (Fig. 1c, d) and macrophages (data not shown) from *Adora2a*^{−/−} mice as compared with cells from *Adora2a*^{+/+} littermates. The function of other G_s-coupled receptors is not affected in *Adora2a*^{−/−} mice, as their ligands have similar effects on cells of both *Adora2a*^{+/+} and *Adora2a*^{−/−} mice (Fig. 1c, d). The Con-A-induced inflammatory damage was so severe in *Adora2a*^{−/−} mice that two out of four male mice in the group died within 8 h, thereby precluding time-course studies. The same stimuli given to control wild-type *Adora2a*^{+/+} mice did not cause death. Even a sub-optimal dose of an inflammatory stimulus (see inset to Fig. 2a) resulted in the death of two out of five male *Adora2a*^{−/−} mice within 48 h, whereas all *Adora2a*^{+/+} controls survived. Low doses of Con A, which caused only minimal or no liver damage in control *Adora2a*^{+/+} mice, were sufficient to induce extensive liver damage in *Adora2a*^{−/−} mice (Fig. 2b). Exacerbated inflammation and tissue damage in *Adora2a*^{−/−} mice could not be explained by increased susceptibility of *Adora2a*^{−/−} hepatocytes to TNF-α, as TNF-α was equally efficient in directly destroying hepatocytes in both *Adora2a*^{−/−} and *Adora2a*^{+/+} mice *in vivo* (Fig. 2c and data not shown). Furthermore, excessive and prolonged pro-inflammatory TNF-α accumulation was observed in the serum of *Adora2a*^{−/−} mice compared with low or undetectable TNF-α levels in *Adora2a*^{+/+} mice (Fig. 2a). IFN-γ was also present at higher concentrations and for a longer duration in *Adora2a*^{−/−} mice, although levels of IL-4 were not different between *Adora2a*^{−/−} and

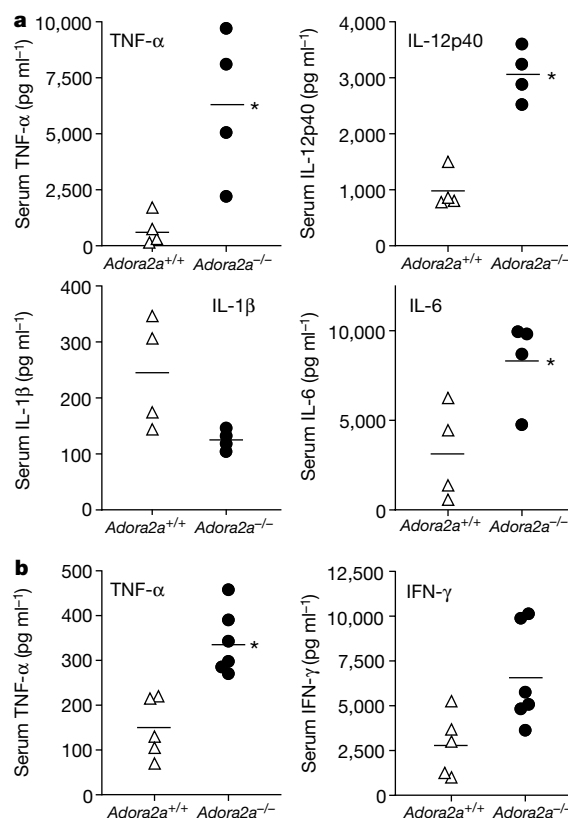


Figure 4 Enhanced accumulation of pro-inflammatory cytokines and tissue damage in A2a-receptor-deficient mice treated with endotoxin. a, LPS (100 μg kg^{−1}) was injected into the dorsal air pouch of *Adora2a*^{−/−} and *Adora2a*^{+/+} mice (*n* = 4), and serum cytokines levels were estimated at different times after LPS injection. Data shown are TNF-α levels at 1 h, IL-12p40 levels at 3 h, IL-1β levels at 3 h and IL-6 levels at 1 h. Asterisk, *P* < 0.05 compared with *Adora2a*^{+/+} mice. b, *Adora2a*^{+/+} and *Adora2a*^{−/−} mice (*n* = 6) were injected intravenously with LPS (5 mg kg^{−1}). Serum cytokine levels after 16 h are shown. Data shown are representative of three separate experiments. Asterisk, *P* < 0.05 compared with *Adora2a*^{+/+} mice.

Adora2a^{+/+} mice (Fig. 2a). Pharmacological inactivation of A2a receptors in *Adora2a*^{+/+} wild-type mice using a selective A2 receptor antagonist, ZM2413856 (refs 26, 28), also resulted in an increased inflammatory tissue damage in *Adora2a*^{-/-} mice (Fig. 3a; see also Supplementary Information).

Tissue-protecting properties of A2 adenosine receptors were further confirmed in other models of inflammatory liver injury and systemic inflammation (Figs 3b, c and 4). Inactivation of A2 adenosine receptors in *Adora2a*^{+/+} mice by an antagonist, ZM241385, exacerbated (Fig. 3b) the T-cell- and TNF- α -dependent acute hepatotoxicity of *Pseudomonas aeruginosa* exotoxin A (PEA)²⁹. The increased liver injury was observed in *Adora2a*^{-/-} mice during chemically induced³⁰ (carbon tetrachloride, CCl₄) acute hepatotoxicity (Fig. 3c). The A2a adenosine receptors are also shown to be crucial in downregulating pro-inflammatory cytokine accumulation and tissue damage (see Supplementary Information) in an *in vivo* septic shock model¹ after subcutaneous (Fig. 4a) and intravenous (Fig. 4b) bacterial endotoxin (lipopolysaccharide, LPS) injection.

The striking phenotype of markedly enhanced liver damage and sepsis in *Adora2a*-deficient mice suggests that no other mechanism of downregulation of inflammation *in vivo* is able to fully compensate for the lack of adenosine receptors in T cells and cells of the innate immune system. The uniqueness of the adenosine receptors may lie in the physiology of accumulation of abundant and ubiquitous purine nucleotides^{5,6} in the local inflammatory environment^{18–22}. Finally, the marked effects of A2a receptor deficiency on functions of immune cells *in vivo* point to the possible role of A2a receptors and other purinergic receptors in the regulation of different stages of immune responses, including antigen presentation, T-cell activation, expansion, survival and memory. □

Methods

Mice

The C57BL/6 A2a-receptor-deficient mice (*Adora2a*^{-/-} N7) were described previously^{25–27}. We determined the A2a receptor genotypes of mice by Southern blot analysis²⁵. Littermates or age-matched wild-type (*Adora2a*^{+/+}) and homozygous (*Adora2a*^{-/-}) mice were used in experiments for better reproducibility of results.

Reagents

Concanavalin A (type IV), D-galactosamine, isoproterenol, PGE₂, LPS (*Escherichia coli* 0111:B4), PEA, CCl₄ and a serum transaminase ALT determination kit were purchased from Sigma. We purchased CGS21680 and ZM241385 from Tocris. Recombinant mouse TNF- α was purchased from Pharmingen.

Measurements of cAMP

Liver mononuclear cells were separated from parenchymal hepatocytes and cell debris by centrifugation using 33% Percoll (see Supplementary Information for detailed methods). Stimulation of cells and the measurement of cAMP levels were executed as described previously²⁶. Liver mononuclear cells (1 $\times$ 10⁵ cells per 200 μ l) were incubated at 37 °C for 30 min in the presence of 10 μ M CGS21680, 1 μ M ZM241385, 100 μ M isoproterenol, or 1 μ M PGE₂. Levels of cAMP were determined using a cAMP enzyme immunoassay kit (Amersham Pharmacia Biotech).

Induction of liver injury

Mice were injected intravenously with Con A (12.5 or 20 mg kg⁻¹) in sterile PBS, and serum samples were taken or mice were killed at indicated time points. Some mice were coinjected intraperitoneally with CGS21680 (2 mg kg⁻¹), ZM241385 (2 mg kg⁻¹), isoproterenol (100 mg kg⁻¹), or PGE₂ (5 mg kg⁻¹) just before treatment with Con A. The magnitude of liver damage was evaluated by serum alanine aminotransferase (ALT) levels and liver tissue histology. Effects of TNF- α on liver injury of *Adora2a*^{+/+} and *Adora2a*^{-/-} mice were compared by injecting mice with a combination of D-galactosamine and TNF- α . D-galactosamine (700 mg per kg) was injected intraperitoneally 30 min before intravenous injection of recombinant mouse TNF- α (4–15 μ g kg⁻¹). After 6 h, mice were killed and the liver damage was evaluated as described.

In studies of PEA-induced (100 μ g kg⁻¹ intravenously) liver injury²⁹, some of the mice were also treated with intraperitoneal injections of ZM241385 before and 12 h after PEA injection. Studies of CCl₄-induced hepatotoxicity³⁰ were conducted after intraperitoneal injections of *Adora2a*^{+/+} and *Adora2a*^{-/-} mice with CCl₄ (0.5 ml kg⁻¹) dissolved in olive oil.

Treatment with endotoxin (LPS)

Endotoxic shock in male *Adora2a*^{-/-} mice and age-matched *Adora2a*^{+/+} mice was induced by intravenous injection of 5 mg kg⁻¹ LPS. At 1 and 16 h after injection, samples of blood

were taken by retro-orbital bleeding. LPS was injected (100 μ g kg⁻¹) into a dorsal air pouch, which was prepared using sterile air essentially as described².

Histological examination and apoptosis assays

Apoptosis assays (the detection of apoptotic cells by *in situ* staining of single-strand breaks in nuclear DNA) were performed by molecular histology. The histological evaluation of tissues after intravenous injections of mice with endotoxin (LPS) was performed by D. Haines. See Supplementary Information for detailed methods.

Measurements of cytokine levels

TNF- α , IFN- γ , IL-4, IL-6, IL-1 β and IL-12p40 levels in the sera and culture supernatant were determined using enzyme-linked immunosorbent assay (ELISA) kits obtained from R&D systems according to the manufacturer's suggestions.

Statistical analysis

Data are expressed as mean $\pm$ s.e.m. Differences between groups were evaluated using Student's *t*-test.

Received 17 September; accepted 8 October 2001.

1. Tarkowski, A. & Wagner, H. Arthritis and sepsis caused by *Staphylococcus aureus*: can the tissue injury be reduced by modulating the host's immune system? *Mol. Med. Today* **4**, 15–18 (1998).
2. Levy, B. D., Clish, C. B., Schmidt, B., Gronert, K. & Serhan, C. N. Lipid mediator class switching during acute inflammation: signals in resolution. *Nature Immunol.* **2**, 612–619 (2001).
3. Gilman, A. G. G proteins: transducers of receptor-generated signals. *Annu. Rev. Biochem.* **56**, 615–649 (1987).
4. Olah, M. E. & Stiles, G. L. Adenosine receptor subtypes: characterization and therapeutic regulation. *Annu. Rev. Pharmacol. Toxicol.* **35**, 581–606 (1995).
5. Fredholm, B. B. et al. Structure and function of adenosine receptors and their genes. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **362**, 364–374 (2000).
6. Burnstock, G. Purine-mediated signaling in pain and visceral perception. *Trends Pharmacol. Sci.* **22**, 182–188 (2001).
7. Linden, J. Molecular approach to adenosine receptors: receptor-mediated mechanisms of tissue protection. *Annu. Rev. Pharmacol. Toxicol.* **41**, 775–787 (2001).
8. Cronstein, B. N. Adenosine, an endogenous anti-inflammatory agent. *J. Appl. Physiol.* **76**, 5–13 (1994).
9. Cronstein, B. N. A novel approach to the development of anti-inflammatory agents: adenosine release at the inflamed sites. *J. Invest. Med.* **43**, 50–57 (1995).
10. Firestein, G. S. et al. Protective effect of an adenosine kinase inhibitor in septic shock. *J. Immunol.* **152**, 5853–5859 (1994).
11. Huang, S., Koshiba, M., Apasov, S. & Sitkovsky, M. Role of A2a adenosine receptor-mediated signaling in inhibition of T cell activation and expansion. *Blood* **90**, 1600–1610 (1997).
12. van der Pouw Kraan, T. C. T. M. et al. Histamine inhibits the production of interleukin-12 through interaction with H2 receptors. *J. Clin. Invest.* **102**, 1866–1873 (1998).
13. Malfait, A.-M. et al. The β 2-adrenergic agonist salbutamol is a potent suppressor of established collagen-induced arthritis: mechanisms of action. *J. Immunol.* **162**, 6278–6283 (1999).
14. Eigler, A. et al. Endogenous adenosine curtails lipopolysaccharide-stimulated tumour necrosis factor synthesis. *Scand. J. Immunol.* **45**, 132–139 (1997).
15. Sullivan, G. W., Sarembock, I. J. & Linden, J. The role of inflammation in vascular diseases. *J. Leukoc. Biol.* **67**, 591–602 (2000).
16. Tiegs, G. Experimental hepatitis and role of cytokines. *Acta Gastroenterol. Belg.* **60**, 176–179 (1997).
17. Kaneko, Y. et al. Augmentation of Valpha14 NK T cell-mediated cytotoxicity by interleukin 4 in an autocrine mechanism resulting in the development of concanavalin A-induced hepatitis. *J. Exp. Med.* **191**, 105–114 (2000).
18. Winn, H. R., Rubio, R. & Berne, R. M. Brain adenosine concentration during hypoxia in rat. *Am. J. Physiol.* **241**, H235–H242 (1981).
19. Van Belle, H., Goossens, F. & Wynants, J. Formation and release of purine catabolites during hypoperfusion, anoxia, and ischemia. *Am. J. Pathol.* **252**, H886–H893 (1987).
20. Rudolph, K. A., Schubert, P., Parkinson, F. E. & Fredholm, B. B. Neuroprotective role of adenosine in cerebral ischemia. *Trends Pharmacol. Sci.* **13**, 439–445 (1992).
21. Marquardt, D. L., Gruber, H. E. & Wasserman, S. I. Adenosine release from stimulated mast-cells. *Proc. Natl Acad. Sci. USA* **81**, 6192–6196 (1984).
22. Filippini, A., Taffs, R. E. & Sitkovsky, M. V. Extracellular ATP in T-lymphocyte activation: possible role in effector functions. *Proc. Natl Acad. Sci. USA* **87**, 8267–8271 (1990).
23. Hoskin, D. W., Reynolds, T. & Blay, J. Adenosine as a possible inhibitor of killer T-cell activation in the microenvironment of solid tumours. *Int. J. Cancer* **59**, 854–855 (1994).
24. Ledent, C. et al. Aggressiveness, hypoalgesia and high blood pressure in mice lacking the adenosine A2a receptor. *Nature* **388**, 674–678 (1997).
25. Chen, J.-F. et al. A2a adenosine receptor deficiency attenuates brain injury induced by transient focal ischemia in mice. *J. Neurosci.* **19**, 9192–9200 (1999).
26. Apasov, S. G. et al. Study of A2a adenosine receptor gene deficient mice reveals that adenosine analogue CGS 21680 possesses no A2a receptor-unrelated lymphotoxicity. *Br. J. Pharmacol.* **131**, 43–50 (2000).
27. Armstrong, J. M. et al. Gene dose effect reveals no Gs protein coupled A2a adenosine receptor reserve in murine T lymphocytes. Studies of cells from A2a receptor gene-deficient mice. *J. Biochem.* **354**, 123–130 (2001).
28. Ji, X.-D. & Jacobson, K. A. Use of the triazolotriazine [3H]ZM 241385 as a radioligand at recombinant human A2b adenosine receptors. *Drug Design Disc.* **16**, 217–226 (1999).
29. Shumann, J., Angermuller, S., Bang, R., Lohoff, M. & Tiegs, G. Acute hepatotoxicity of *Pseudomonas aeruginosa* exotoxin A in mice depends on T cells and TNF. *J. Immunol.* **161**, 5745–5754 (1998).
30. Luster, M. I. et al. Role of inflammation in chemical-induced hepatotoxicity. *Toxicol. Lett.* **120**, 317–321 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Fan Chen, S. Fink and M. A. Schwartzschild for providing A2a receptor gene deficient mice; W. E. Paul, R. N. Germain, S. Apasov, P. Smith and P. M. Murphy for support, discussions and help; and J. Kinsel and S. Starnes for help in preparation of the manuscript.

Correspondence and requests for materials should be addressed to M.S. (e-mail: mvsitkov@helix.nih.gov).

MIF regulates innate immune responses through modulation of Toll-like receptor 4

Thierry Roger^{*}, John David[†], Michel P. Glauser^{*} & Thierry Calandra^{*}

^{*} Division of Infectious Diseases, Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, rue du Bugnon 46, CH-1011 Lausanne, Switzerland

[†] Department of Immunology and Infectious Diseases, Harvard School of Public Health, 665 Huntington Ave., Boston, Massachusetts 02115, USA

Macrophages are pivotal effector cells of the innate immune system, which is vital for recognizing and eliminating invasive microbial pathogens^{1,2}. When microbial products bind to pathogen-recognition receptors, macrophages become activated and release a broad array of cytokines³ that orchestrate the host innate and adaptive immune responses. Initially identified as a T-cell cytokine^{4,5}, macrophage migration inhibitory factor (MIF) is also a macrophage cytokine and an important mediator of inflammation and sepsis^{6–12}. Here we report that MIF is an essential regulator of macrophage responses to endotoxin (lipopolysaccharide) and Gram-negative bacteria. Compared with wild-type cells, MIF-deficient macrophages are hyporesponsive to lipopolysaccharide and Gram-negative bacteria, as shown by a profound reduction in the activity of NF- κ B and the production of tumour-necrosis factor- α . This reduction is due to a downregulation of Toll-like receptor 4 (TLR4), the signal-transducing molecule of the lipopolysaccharide receptor complex, and is associated with decreased activity of transcription factor PU.1, which is required for optimal expression of the *Tlr4* gene in myeloid cells. These findings identify an important role for MIF in innate immunity and provide a molecular basis for the resistance of MIF-deficient mice to endotoxic shock.

Monocytes and macrophages constitutively express copious amounts of MIF; such expression is an unusual feature for a cytokine⁷. Given that MIF has been shown to exert intracellular activity¹³, we reasoned that reducing the endogenous content of MIF in macrophages might help to unravel the function of MIF in the innate immune system.

RAW 264.7 mouse macrophages were stably transfected with a plasmid encoding an antisense MIF messenger RNA. Two clones (designated as AS 2.8 and AS 2.23) exhibiting a marked reduction in MIF protein were compared with control macrophages transfected with the empty plasmid (Fig. 1a). In contrast to control macrophages, antisense MIF macrophages were hyporesponsive to endotoxin (*Escherichia coli* O111:B4 lipopolysaccharide; LPS) stimulation, as shown by a profound reduction in tumour-necrosis factor- α (TNF- α) and interleukin 6 (IL-6) (Fig. 1b; and data not shown). Similar results were obtained when antisense MIF macrophages were exposed to heat-killed *E. coli*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*—the three most frequent causes of Gram-negative sepsis in humans (Fig. 1c).

In contrast, control and antisense MIF macrophages responded equally well to stimulation with heat-killed Gram-positive bacteria (*Staphylococcus aureus*, group A streptococci), peptidoglycan or the yeast particle zymosan, and to stimulation with phorbol 12-myristate 13-acetate (PMA) plus calcium ionophore (Fig. 1d; and data not shown). Thus, antisense MIF macrophages displayed a defect in response to LPS and Gram-negative bacteria, but not to other microbial stimuli, indicating that MIF might have a role in the signalling pathways activated by LPS in the macrophage.

To explore the molecular mechanism by which MIF regulates the responses of macrophages to LPS and Gram-negative bacteria, we examined the intracellular events that control expression of the TNF- α gene. The DNA-binding activity of nuclear factor kappa B (NF- κ B), the luciferase reporter activity driven by NF- κ B, and the concentrations of TNF- α mRNA (Fig. 1e; and Supplementary Information) were all markedly decreased in antisense MIF macrophages stimulated with LPS. These results indicate that reducing the endogenous concentrations of MIF in the macrophage may impede LPS signal transduction at a point positioned between the binding of LPS to its receptor complex and the activation of NF- κ B.

Recognition of LPS and Gram-negative bacteria by the host requires cooperative interplay between the LPS-binding protein (LBP)¹⁴, CD14 (ref. 15) and Toll-like receptor 4 (TLR4)¹⁶. LBP binds and transfers LPS-containing particles to a receptor complex

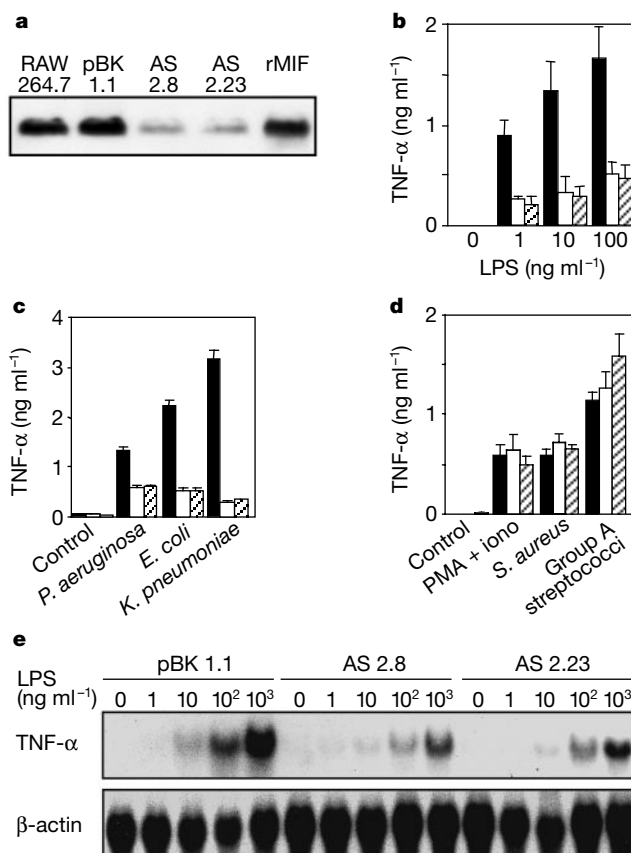


Figure 1 Hyporesponsiveness of antisense MIF macrophages to activation by LPS and Gram-negative bacteria. **a**, Western blot of intracellular MIF content of RAW 264.7 macrophages stably transfected with an empty plasmid (pBK 1.1) or with an antisense MIF cDNA plasmid (AS 2.8 and AS 2.23). rMIF, recombinant MIF. **b–d**, TNF- α production by pBK 1.1 (black bars), AS 2.8 (white bars) and AS 2.23 (hatched bars) macrophages stimulated with LPS, 10⁸ c.f.u. of heat-killed bacteria, or PMA plus ionomycin (PMA+iono). Data are the mean $\pm$ s.d. of triplicate samples from 1 experiment and are representative of 3–9 independent experiments. **e**, Northern blots of TNF- α and β -actin mRNA in pBK 1.1, AS 2.8 and AS 2.23 macrophages stimulated for 2 h with LPS. Data are representative of four separate experiments.

composed of CD14, a ligand-binding GPI-anchored protein, and TLR4, the molecule that transduces the LPS signal. TLR4, the first identified human homologue of the *Drosophila* Toll receptor, was originally shown to activate NF- κ B and to induce the expression of pro-inflammatory cytokines by human monocytic cells¹⁷. The defective LPS signalling of C3H/HeJ and C57Bl/10ScCr mice has been linked to missense and null mutations, respectively, of the *Tlr4* gene¹⁶. This crucial observation and a series of studies in mice and humans have shown unequivocally that TLR4 is an essential component of the LPS receptor complex^{18–21}. We therefore investigated whether LBP, CD14 or TLR4 is implicated in the impaired response of antisense MIF macrophages to LPS.

Control and antisense macrophages expressed comparable membrane-bound levels of CD14 and bound complexes of fluorescein isothiocyanate (FITC)-labelled LPS and LBP equally well (see Supplementary Information). In contrast to CD14, however, the

baseline levels of TLR4 mRNA (Fig. 2a) and protein (Fig. 2b) were considerably reduced in antisense MIF macrophages. Reduction of TLR4 expression and TNF- α production were also decreased significantly in macrophages transduced with an antisense adenovirus system or in cells treated with anti-MIF antibodies (data not shown). Messenger RNA concentrations of MD-2, a secreted protein that associates with the extracellular domain of TLR4 and augments the cell response to LPS²², were similar in control and antisense MIF macrophages (Fig. 2a). But mRNA expression of TLR2, a member of the TLR family that mediates host responses to components of Gram-positive bacteria, mycobacteria and yeast²³, was preserved in antisense MIF macrophages (Fig. 3a). Conservation of TLR2 expression is congruous with the observation that MIF-deficient macrophages responded normally to Gram-positive bacteria, peptidoglycan and zymosan (Figs 1d and 3e).

To verify that the downregulation of TLR4 accounts for the abnormal response of antisense MIF macrophages to LPS stimulation, we transfected cells with an expression vector encoding a mouse TLR4 complementary DNA. Adding back mouse TLR4 corrected the defective response to LPS, as shown by the restoration of NF- κ B activity and TNF- α production (Fig. 2c, d; and Supplementary information). Similar results were obtained when antisense MIF macrophages were transfected with an expression vector

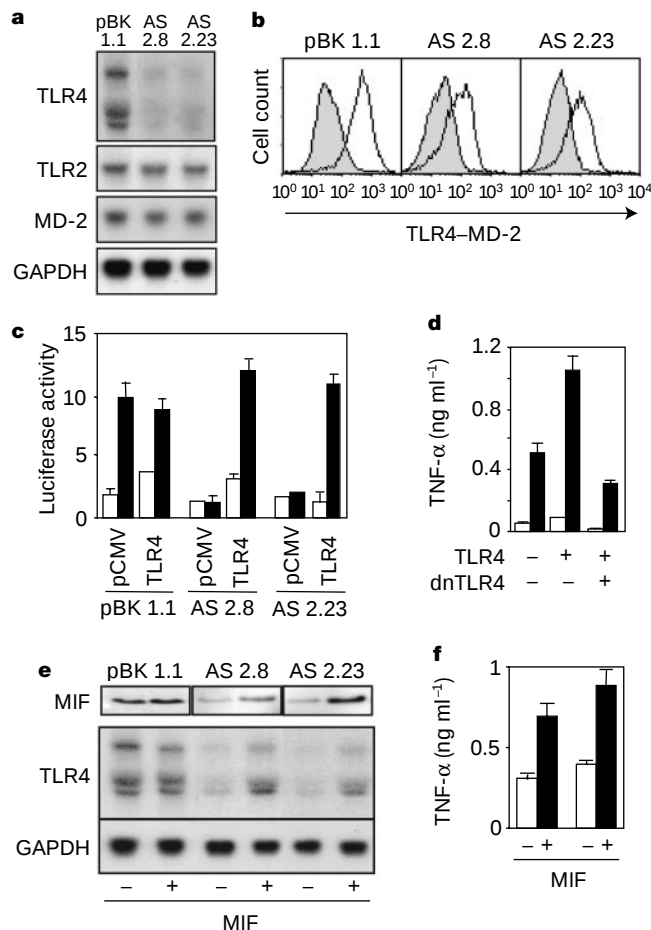


Figure 2 MIF regulates macrophage responses to LPS by modulating the expression of TLR4. **a**, Northern blots of basal TLR4, TLR2, MD-2 and GAPDH mRNA in pBK 1.1, AS 2.8 and AS 2.23 RAW 264.7 macrophages. **b**, TLR4–MD-2 expression on pBK 1.1, AS 2.8 and AS 2.23 macrophages analysed by flow cytometry. Background staining is shown in grey. **c**, Transfection of antisense MIF macrophages with mouse TLR4–pCMV restores responsiveness to LPS, as assessed by measuring NF- κ B-driven luciferase activities 6 h after stimulation with (black bars) or without (white bars) LPS (100 ng ml⁻¹). **d**, Dominant-negative TLR4 abolished TNF- α production by antisense MIF macrophages transfected with TLR4 and stimulated for 4 h with (black bars) or without (white bars) LPS (100 ng ml⁻¹). **e**, **f**, Raising the MIF content of antisense MIF macrophages restores TLR4 expression and responsiveness to LPS. pBK 1.1, AS 2.8 and AS 2.23 macrophages were transfected with an empty plasmid (–) or with a MIF expression plasmid (+). **e**, Western blot of intracellular MIF and northern blots of TLR4 and GAPDH mRNA. **f**, TNF- α production by AS 2.8 and AS 2.23 stimulated for 4 h with (black bars) or without (white bars) LPS (100 ng ml⁻¹).

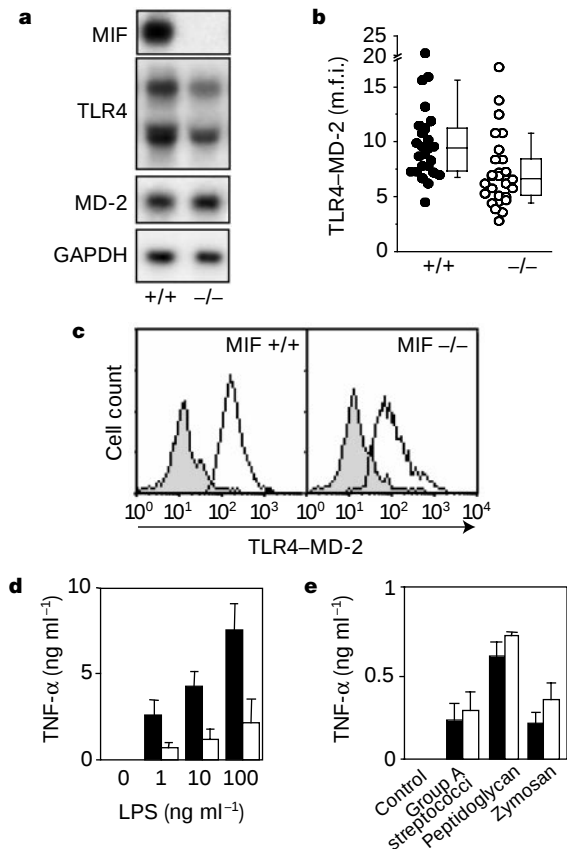


Figure 3 Decreased basal TLR4 expression and LPS-induced TNF- α production by MIF knockout (MIF^{-/-}) macrophages. **a**, Northern blots of basal MIF, TLR4, MD-2 and GAPDH mRNA of MIF^{+/+} and MIF^{-/-} macrophages. **b**, Flow cytometry analysis of TLR4–MD-2 expression of thioglycollate-elicited peritoneal macrophages isolated from MIF^{+/+} ($n = 25$) and MIF^{-/-} ($n = 27$) mice¹¹ ($P = 0.002$). m.f.i., mean fluorescence intensity. Box plots show the 10th, 25th, 50th, 75th and 90th percentiles, respectively. **c**, Representative histograms of TLR4–MD-2 expression in MIF^{+/+} and MIF^{-/-} macrophages. Background staining is shown in grey. **d**, **e**, TNF- α production by MIF^{+/+} (black bars) and MIF^{-/-} (white bars) macrophages stimulated with LPS, 10⁸ colony forming units (c.f.u.) of heat-killed bacteria, peptidoglycan (100 μ g ml⁻¹) or zymosan (2 $\times$ 10⁶). Data are the mean $\pm$ s.d. of two independent experiments with cells isolated from a total of four mice per group.

encoding human TLR4, whereas cells transfected with human TLR1, TLR2 or TLR3 remained unresponsive to LPS (data not shown). Moreover, the restoration of normal LPS signalling in antisense MIF macrophages transfected with mouse TLR4 was inhibited by co-expression of a dominant-negative mutant of mouse TLR4 (Fig. 2d; and Supplementary Information). Alternatively, the endogenous MIF protein content could be reconstituted by transiently transfecting antisense MIF macrophages with an expression vector encoding a mouse MIF cDNA. Raising the intracellular MIF content of antisense MIF macrophages to concentrations similar to those in control macrophages restored the basal expression of TLR4 mRNA and the production of TNF- α in cells stimulated with LPS (Fig. 2e, f). Together, these results indicate that a reduction in endogenous MIF may impair the ability of macrophages to respond to LPS and Gram-negative bacteria through the downregulation of TLR4.

We examined the importance of MIF in host responses to endotoxin *in vivo* using macrophages from MIF knockout (MIF^{-/-}) mice¹¹. Thioglycollate-elicited peritoneal macrophages were isolated from MIF^{-/-} and MIF^{+/+} mice. The expression of TLR4 mRNA (Fig. 3a) and protein (Fig. 3b, c) was significantly downregulated in MIF^{-/-} macrophages as compared with control MIF^{+/+} macrophages. MIF^{-/-} macrophages were also hyporesponsive to stimulation with LPS, as manifested by a marked reduction of TNF- α output over a broad range of LPS concentrations (Fig. 3d). By contrast, MIF^{-/-} and MIF^{+/+} macrophages released similar quantities of TNF- α after stimulation with heat-killed Gram-positive bacteria, peptidoglycan or the yeast particle zymosan (Fig. 3e), indicating that the defect in the response of MIF^{-/-} macrophages was restricted to LPS and Gram-negative bacteria.

We next investigated the molecular mechanism by which MIF modulates the expression of TLR4. Given that levels of TLR4 mRNA are decreased in MIF-deficient macrophages, we first analysed the transcriptional regulation of the *Tlr4* gene in RAW 264.7 macro-

phages transiently transfected with deletion constructs of the promoter region (-2,652 to +223 base pairs (bp)) of the mouse *Tlr4* gene cloned in a luciferase reporter plasmid. A region located between nucleotides -518 and -102 positively regulated luciferase activity driven by the *Tlr4* promoter (Fig. 4a). It comprised two binding sites (TTCCTCT), located at positions -292 to -286 and -115 to -109, for the transcription factor PU.1. This transcription factor has an important role in the transcription of the human *Tlr4* gene in myeloid cells²⁴. We found that disrupting the two PU.1-binding sites by site-directed mutagenesis either completely abolished (distal site) or strongly reduced (proximal site) basal luciferase activity in RAW 264.7 macrophages (data not shown), indicating that PU.1 has a crucial role in transcription of the murine *Tlr4* gene.

We transfected control and MIF-deficient RAW 264.7 macrophages with the -518 *Tlr4* promoter luciferase construct to compare the *Tlr4* promoter activities of both cell types. *Tlr4* promoter activity was decreased by 60% in MIF-deficient macrophages as compared with control cells (Fig. 4b). Similar results were obtained when cells were transfected with a trimeric major histocompatibility complex (MHC)-derived PU.1 luciferase reporter construct (Fig. 4c)²⁵. Finally, basal PU.1 DNA-binding activity, as assessed by electromobility shift assay (EMSA), was markedly decreased in nuclear extracts of MIF-deficient macrophages (Fig. 4d). Together, these results indicate that reduced expression of MIF in macrophages impairs basal PU.1 DNA-binding activity and PU.1-dependent transcriptional activity of the mouse *Tlr4* gene, resulting in reduced expression of TLR4 protein and a defective innate immune response to LPS and Gram-negative bacteria.

These results indicate that MIF may be important in the recognition of LPS and Gram-negative bacteria by cells of the innate immune system. Moreover, they also provide a rationale for the intriguing observation that MIF is expressed constitutively by macrophages and by tissues in direct contact with the

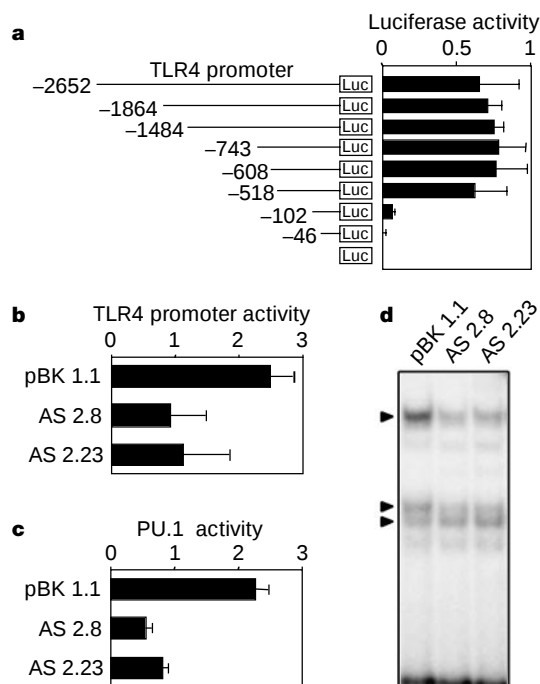


Figure 4 Basal TLR4 promoter activity and PU.1 DNA-binding activity are reduced in antisense MIF RAW 264.7 macrophages. **a**, Activity of deletion constructs of the mouse *Tlr4* promoter cloned into a luciferase reporter plasmid and transfected into RAW 264.7 macrophages. **b**, *Tlr4* promoter activity of pBK 1.1, AS 2.8 and AS 2.23 macrophages transfected with the -518 TLR4 luciferase construct. **c**, PU.1 activity of pBK1.1, AS 2.8

and AS 2.23 macrophages transfected with a trimeric MHC-derived PU.1 luciferase reporter construct. Data are the mean $\pm$ s.d. of 2–5 independent experiments (**a–c**). **d**, Electrophoretic mobility shift assay showing reduced basal PU.1 DNA-binding activity in antisense MIF macrophages. PU.1-specific complexes are marked with arrowheads.

environment^{7,26}. Recognition of invasive microbial pathogens is an essential function of innate immunity. By upregulating the basal expression of TLR4 in the macrophage, MIF promotes the recognition of endotoxin-containing particles and Gram-negative bacteria by the innate immune system. Thereby, MIF enhances the production of inflammatory cytokines such as TNF- α , and facilitates the initiation of the host defence response. Therefore, MIF enables cells that are at the frontline of antimicrobial host defences, such as the macrophage, to respond quickly to hostile Gram-negative bacteria. Our findings identify a crucial role for MIF in innate immunity and provide a rationale for an anti-MIF strategy of treatment in people with Gram-negative septic shock. □

Methods

Reagents and bacteria

Escherichia coli O111:B4 LPS, FITC-labelled *Serratia marcescens* LPS, PMA and calcium ionophore A23187 were obtained from Sigma. G418 was from Brunschwig. *S. aureus* peptidoglycan was from Fluka, zymosan was from Molecular Probes. Peptidoglycan and zymosan contained less than 10 pg of endotoxin per microgram of material. *E. coli* O111 and O18, *P. aeruginosa*, *K. pneumoniae*, *S. aureus* and group A streptococci (*S. pyogenes*) were grown overnight, enumerated, heat-killed, divided into aliquots and frozen at -20°C .

Cells, constructs and transfections

We cultured mouse RAW 264.7 macrophages in RPMI 1640 medium containing 2 mM glutamine, 10% heat-inactivated fetal calf serum (Seromed), 100 IU ml⁻¹ penicillin and 100 $\mu\text{g ml}^{-1}$ streptomycin. Thioglycollate-elicited peritoneal macrophages were obtained from MIF knockout¹¹ and MIF wild-type mice as described⁷. Murine MIF cDNA was cloned in a 3'-5' orientation in the cytomegalovirus promoter-containing expression vector pBK (Stratagene)²⁷. RAW 264.7 macrophages were transfected with the parental pBK vector or the recombinant antisense MIF vector using DOTAP (Roche Diagnostics). We selected stable transfected cell lines using 250 $\mu\text{g ml}^{-1}$ of G418 (Gibco BRL) and clones obtained by limited dilution. The MIF content of stable antisense MIF macrophages was analysed by western blotting and enzyme-linked immuno-absorbent assay as described^{7,8}. For add-back experiments, antisense MIF RAW 264.7 macrophages were transiently transfected with 2 μg of mouse MIF expression construct obtained by cloning the coding sequence of mouse MIF into the pBK vector (Stratagene). The coding regions of mouse TLR2 and TLR4 (amino acids 20–784 for TLR2, and 25–835 for TLR4) were amplified by polymerase chain reaction (PCR) from RAW 264.7 macrophages, cloned into the pGEM-T easy vector (Promega), sequenced and subcloned in the pFlag-CMV expression plasmid (Sigma). We verified proper expression of the constructs by western blot using anti-Flag antibodies (Sigma). A truncated form of mouse TLR4 (amino acids 25–770) with a deletion of the carboxy-terminal 65 amino acids of the intracytoplasmic domain of the molecule was obtained by digesting the TLR4 DNA insert at the EcoNI site and at the BamHI site present in the 3' polylinker of pFlag-CMV.

RNA analysis

Total RNA was extracted from control and antisense MIF RAW 264.7 macrophages and from MIF^{+/+} and MIF^{-/-} peritoneal macrophages. Expression of mouse TNF- α , CD14, TLR2, TLR4, MD-2, β -actin and GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA was assessed by northern blot. We quantified gene-specific mRNA signals using an Instant Imager 2024 (Packard).

Cytokine measurements

The concentrations of TNF- α and IL-6 were determined using WEHI164 clone 13 mouse fibrosarcoma cells (TNF- α) and 7TD1 IL-6-dependent mouse-mouse hybridoma cells (IL-6), as described¹².

Electromobility shift assay

Nuclear extracts of control or LPS-stimulated RAW 264.7 or antisense MIF macrophages were analysed by EMSA using a consensus NF- κB probe (Santa Cruz) and a PU.1 probe derived from the sequence of the distal PU.1-binding site of the mouse *Tlr4* promoter.

Mouse TLR4 promoter

A 3-kb fragment corresponding to region -2715/+223 from the mouse *Tlr4* gene²⁴ was amplified by PCR from genomic DNA of a C57BL/6 mouse and cloned into the pGL3-basic vector (Promega). We obtained deletion constructs of the *Tlr4* promoter using exonuclease III or restriction endonucleases. Sequence analysis of the -518 to -102 bp region of the *Tlr4* promoter revealed the presence of two conserved PU.1-binding sites at positions -292 to -286 and -115 to -109. Mutations confirmed by sequencing of the two PU.1-binding sites (from TTCCTCT to TAGTCT) were obtained by PCR using the QuikChange site-directed mutagenesis kit (Stratagene).

NF- κB , TLR4 promoter and PU.1 luciferase assays

Control and antisense MIF RAW 264.7 macrophages were incubated in six-well tissue-culture plates and transiently transfected on the following day using Eugene6 transfection

reagent (Roche Diagnostics). We transfected cells with 2 μg of a multimeric NF- κB pGL2 luciferase vector and 0.4–2 μg of the other expression constructs, or with 2 μg of *Tlr4* promoter pGL3 (ref. 24) or an MHC-derived PU.1 (ref. 25) luciferase vector and 0.04–0.2 μg of the *Renilla* pRL-TK vector (Promega). Twenty-four hours after transfection, cells were stimulated for 6 h with LPS (100 ng ml⁻¹) and luciferase and *Renilla* luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega). Results are expressed as relative luciferase activity (the ratio of luciferase to *Renilla* luciferase activity).

Flow cytometric analysis

We assessed the binding of FITC-labelled *S. marcescens* LPS-LBP complexes onto RAW 264.7 macrophages by flow cytometry as described²⁸. After blocking Fc receptors with 2.4G2 hybridoma supernatant, expression of membrane-bound CD14 and TLR4-MD-2 complex was evaluated by first incubating cells with a rat anti-mouse CD14 (ref. 29) or TLR4-MD-2 (ref. 30) or an isotype-matched control monoclonal antibody and then with a FITC-conjugated goat anti-rat antibody (Biosource International). We counterstained macrophages with a phycoerythrin-conjugated anti-Mac1 monoclonal antibody (Pharmingen). The amounts of TLR4-MD-2 on MIF^{+/+} and MIF^{-/-} macrophages are expressed as a fold increase of the mean fluorescence intensity over background.

Received 30 August; accepted 16 October 2001.

- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. A. Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318 (1999).
- Aderem, A. & Underhill, D. M. Mechanisms of phagocytosis in macrophages. *Annu. Rev. Immunol.* **17**, 593–623 (1999).
- Dinarello, C. A. Proinflammatory cytokines. *Chest* **118**, 503–508 (2000).
- David, J. Delayed hypersensitivity *in vitro*: its mediation by cell-free substances formed by lymphoid cell-antigen interaction. *Proc. Natl Acad. Sci. USA* **56**, 72–77 (1966).
- Bloom, B. R. & Bennett, B. Mechanism of a reaction *in vitro* associated with delayed-type hypersensitivity. *Science* **153**, 80–82 (1966).
- Bernhagen, J. *et al.* MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia. *Nature* **365**, 756–759 (1993).
- Calandra, T., Bernhagen, J., Mitchell, R. A. & Bucala, R. The macrophage is an important and previously unrecognized source of macrophage migration inhibitory factor. *J. Exp. Med.* **179**, 1895–1902 (1994).
- Calandra, T. *et al.* MIF as a glucocorticoid-induced modulator of cytokine production. *Nature* **377**, 68–71 (1995).
- Bacher, M. *et al.* An essential regulatory role for macrophage migration inhibitory factor in T-cell activation. *Proc. Natl Acad. Sci. USA* **93**, 7849–7854 (1996).
- Calandra, T., Spiegel, L. A., Metz, C. N. & Bucala, R. Macrophage migration inhibitory factor is a critical mediator of the activation of immune cells by exotoxins of Gram-positive bacteria. *Proc. Natl Acad. Sci. USA* **95**, 11383–11388 (1998).
- Bozza, M. *et al.* Targeted disruption of migration inhibitory factor gene reveals its critical role in sepsis. *J. Exp. Med.* **189**, 341–346 (1999).
- Calandra, T. *et al.* Protection from septic shock by neutralization of macrophage migration inhibitory factor. *Nature Med.* **6**, 164–170 (2000).
- Kleemann, R. *et al.* Intracellular action of the cytokine MIF to modulate AP-1 activity and the cell cycle through Jab1. *Nature* **408**, 211–216 (2000).
- Schumann, R. R. *et al.* Structure and function of lipopolysaccharide binding protein. *Science* **249**, 1429–1431 (1990).
- Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* **249**, 1431–1433 (1990).
- Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* **282**, 2085–2088 (1998).
- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J. & Gusovsky, F. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J. Biol. Chem.* **274**, 10689–10692 (1999).
- Hoshino, K. *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J. Immunol.* **162**, 3749–3752 (1999).
- Heine, H. *et al.* Cutting edge: cells that carry a null allele for toll-like receptor 2 are capable of responding to endotoxin. *J. Immunol.* **162**, 6971–6975 (1999).
- Arbour, N. C. *et al.* TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nature Genet.* **25**, 187–191 (2000).
- Shimazu, R. *et al.* MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med.* **189**, 1777–1782 (1999).
- Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature Immunol.* **2**, 675–680 (2001).
- Rehli, M. *et al.* PU.1 and interferon consensus sequence-binding protein regulate the myeloid expression of the human Toll-like receptor 4 gene. *J. Biol. Chem.* **275**, 9773–9781 (2000).
- Matsumura, I. *et al.* Biologic significance of GATA-1 activities in Ras-mediated megakaryocytic differentiation of hematopoietic cell lines. *Blood* **96**, 2440–2450 (2000).
- Bacher, M. *et al.* Migration inhibitory factor expression in experimentally induced endotoxemia. *Am. J. Pathol.* **150**, 235–246 (1997).
- Waeber, G. *et al.* Insulin secretion is regulated by the glucose-dependent production of islet b-cell macrophage migration inhibitory factor. *Proc. Natl Acad. Sci. USA* **94**, 4782–4787 (1997).
- Le Roy, D. *et al.* Monoclonal antibodies to murine lipopolysaccharide (LPS)-binding protein (LBP) protect mice from lethal endotoxemia by blocking either the binding of LPS to LBP or the presentation of LPS/LBP complexes to CD14. *J. Immunol.* **162**, 7454–7460 (1999).
- Adachi, Y. *et al.* Inhibition by a CD14 monoclonal antibody of lipopolysaccharide binding to murine macrophages. *J. Endotoxin Res.* **5**, 139–146 (1999).
- Akashi, S. *et al.* Cutting edge: cell surface expression and lipopolysaccharide signaling via the toll-like receptor 4-MD-2 complex on mouse peritoneal macrophages. *J. Immunol.* **164**, 3471–3475 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank M. Knap and D. LeRoy for technical assistance, and I. Miconnet, P. Majcherczyk and C. Froidevaux for critically reading the manuscript. We are grateful to G. Waeber and to I. Matsumura for their gifts of antisense MIF mRNA and the MHC-PU.1 luciferase construct, respectively. We thank T. Yadomae and K. Miyake for providing monoclonal antibodies to mouse CD14 and TLR4-MD-2, respectively. This work was supported by a grant from the Swiss National Science Foundation to T.C. T.C. and M.P.G. are recipients of career awards from the Swiss National Science Foundation, the Leenards Foundation and the Bristol-Myers Squibb Foundation.

Correspondence and requests for materials should be addressed to T.C. (e-mail: Thierry.Calandra@chuv.hospvd.ch).

Selectivity of chromatin-remodelling cofactors for ligand-activated transcription

Bryan Lemon, Carla Inouye, David S. King & Robert Tjian

Department of Molecular and Cell Biology, Howard Hughes Medical Institute, 401 Barker Hall, University of California, Berkeley, California 94720-3204, USA

An array of regulatory protein and multi-subunit cofactors has been identified¹ that directs eukaryotic gene transcription. However, establishing the specific functions of various related cofactors has been difficult owing to the limitations inherent in assaying transcription in animals and cells indirectly. Here we describe, using an integrated chromatin-dependent reconstituted transcription reaction, the purification and identification of a multi-subunit cofactor (PBAF)² that is necessary for ligand-dependent transactivation by nuclear hormone receptors. A highly related cofactor, human SWI/SNF³, and the ISWI-containing chromatin-remodelling complex ACF⁴ both fail to potentiate transcription. We also show that transcriptional activation mediated by nuclear hormone receptors requires TATA-binding protein (TBP)-associated factors (TAFs) as well as the multi-subunit cofactors ARC⁵/CRSP⁶. These studies demonstrate functional selectivity amongst highly related complexes involved in gene regulation and help define a more complete set of factors and cofactors required to activate transcription.

The transcription of select genes is directed by sequence-specific DNA-binding proteins like Sp1 and members of the steroid and nuclear hormone receptor superfamily such as the vitamin D3 receptor (VDR), retinoid X receptors (RXRs) and peroxisome proliferator-activated receptors (PPARs). The diversity of transcriptional cofactors that mediate gene activation or repression by such promoter-specific proteins includes specialized protein complexes that are thought to help RNA polymerase navigate through chromatin^{1,7}. To examine how specificity is achieved, we sought to reconstruct an integrated transcription reaction with chromatin templates that could recapitulate the tight promoter control and ligand dependence of nuclear receptor transactivation observed *in vivo*. Our *in vitro* transcription assay includes recombinant human general transcription factors TFIIA, -B, -E and -F together with highly purified human RNA Pol II, and immunopurified TFIID and TFIIF. Activated transcription with this biochemical system further requires the coactivators CRSP⁶ and ARC⁵. This transcription system was used to search for potentially important cofactors which may inadvertently have been present but overlooked in previous reactions^{5,8}.

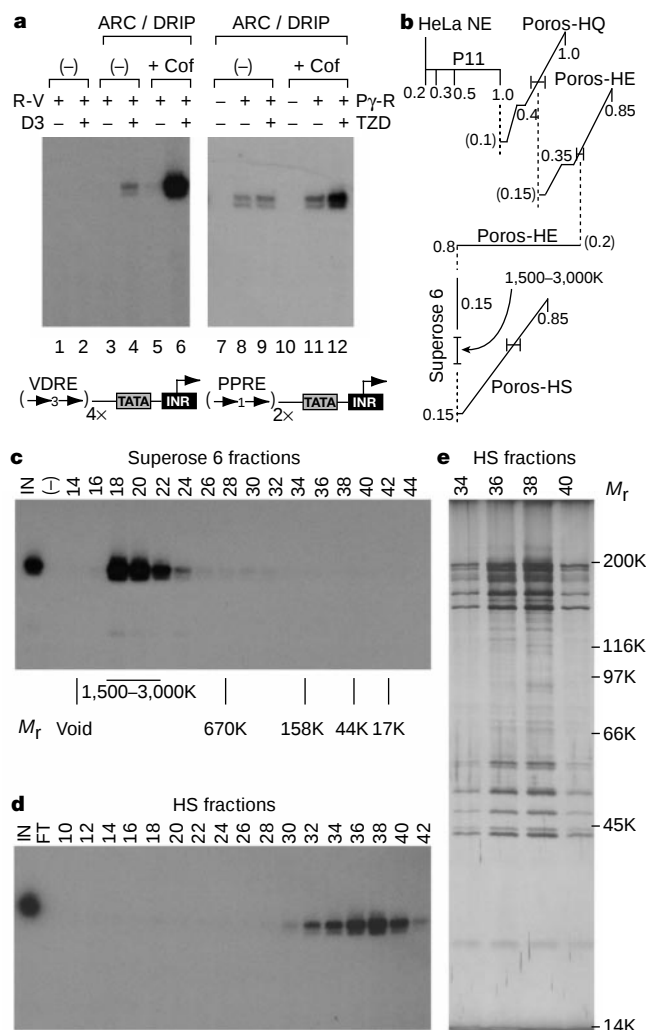


Figure 1 Purification of a cofactor required for activated transcription on chromatin templates *in vitro*. **a**, New activity for ligand-dependent nuclear hormone receptor activated transcription. Reconstituted *in vitro* transcription reactions (Methods) supplemented with RXR + VDR (R-V, lanes 1–6) and 100 nM 1,25(OH)₂-vitamin D₃ (D3, lanes 2, 4, 6) or PPARγ + RXR (Pγ-R, lanes 8, 9, 11, 12) and 1 μM rosiglitazone (TZD, lanes 9, 12), together with affinity-purified ARC/DRIP (lanes 3–12) plus a cofactor fraction (Cof, lanes 5, 6, 10–12) and programmed with pre-assembled chromatin templates (bottom, see Methods). INR, initiator. **b**, Chromatography scheme for purification of cofactor activity from nuclear extracts (HeLa NE). Numbers indicate KCl molarity, parentheses denote dialysis or dilution, and brackets designate pools. Chromatography columns include the cation exchangers phosphocellulose (P11), heparin (Poros-HE), and Poros-HS, the anion exchanger Poros-HQ, and gel filtration medium Superose 6. **c**, Cofactor activity migrates as a large complex. Input (IN), buffer (–), and Superose 6 fractions (top) assayed in the presence of RXR + VDR and ligand as in **a**. Mobilities of peak activity (1,500–3,000K) and gel filtration protein standards (BioRad) are depicted (bottom). **d**, Nuclear receptor-activated transcription dependent on a highly purified cofactor. Input (IN), flowthrough (FT), and Poros-HS fractions (top) assayed as in **c**. **e**, Silver-stained SDS 6–15% polyacrylamide gradient gel of active Poros-HS fractions (top) equivalent to 50 transcription reactions in **d**.

We unmasked an activity in human nuclear fractions that strongly potentiated ligand-dependent transcription mediated by affinity-purified ARC⁵/DRIP⁸ from chromatin templates (Fig. 1a, lanes 1–6) using a VDR-dependent transcription assay. Interestingly, PPARγ–RXR responsiveness to the thiazolidinedione (TZD) ligand rosiglitazone was also strictly dependent upon this cofactor fraction (lanes 7–12). This putative new transcriptional cofactor was apparently distinct from CBP⁹, PC4¹⁰, topoisomerase I¹¹, and

CRSP⁶, as these other cofactors failed to complement the observed activity (data not shown). We used the reconstituted transcription system together with affinity-purified ARC and CBP supplemented with a CRSP-containing fraction, along with nuclear receptors, ligand, and a chromatin template, to purify this cofactor over five successive chromatographic columns (Fig. 1b) resulting in an 8,000 to 15,000-fold increase in specific activity. Strong dependence on cofactor activity (>100-fold) was observed in transcription reactions using the most purified fractions (Fig. 1c, d). Because the activity migrated with an apparent relative molecular mass of 2×10^6 ($M_r = 2,000K$) during size-exclusion chromatography (Fig. 1c), we surmised that the cofactor is a large multi-protein complex. Accordingly, SDS-polyacrylamide gel electrophoresis (PAGE) and silver staining of the most purified fractions revealed a distinct pattern of polypeptides of M_r 40K–200K that eluted with the cofactor activity (Fig. 1e).

To identify the components of this cofactor complex, peak fractions from Poros-HS ion exchange chromatography were pooled and separated by SDS-PAGE. Tryptic digests of ten individually excised bands analysed by mass spectrometry revealed the cofactor to be a BRG1-associated factor (BAF) complex¹². Detailed comparison of several peptide sequences determined that this BAF complex is highly related or identical to human SWI/SNF-B, recently renamed polybromo- and BAF-containing complex (PBAF)². We obtained from 3 to 19 peptide mass matches to each subunit of PBAF (Fig. 2a). Nine peptide masses corresponded uniquely to BRG1/SNF2 β ¹³ and not to the highly related polypeptide hBRM/SNF2 α ¹⁴. We carried out western blot analysis using antibodies directed against known SWI/SNF and PBAF subunits to corroborate the mass spectrometry data. Several BAFs were confirmed to be present (Fig. 2b) but hBRM was not detected in the purified complex (data not shown). Although it is unclear whether β - or γ -actin and β -tubulin might contribute to chromatin-remodelling functions involved in transcription, their presence (Fig. 2a) is consistent with observations that PBAF can be localized to kinetochores of mitotic chromosomes² and that human SWI/SNF components associate with the nuclear matrix¹⁵.

Although the human SWI/SNF-A and -B (PBAF) complexes share many common subunits^{7,12}, one component unique to PBAF is BAF180/polybromo, which contains six bromodomains². The bromodomain structural motif was recently shown to bind selectively to acetylated histone tails, which could be important for targeting cofactor complexes to chromatin¹⁶. An antibody directed against BAF180 was reported to immunoprecipitate the PBAF complex specifically². We therefore asked if the transcriptional cofactor activity would be retained by anti-BAF180 affinity-purified complexes obtained from the most purified cofactor preparations (see Supplementary Information). Importantly, beads containing anti-BAF180 affinity-purified PBAF stimulated transcription activity (Fig. 2c, lanes 9, 10). PBAF complexes retained on beads (~10-fold less than the input) exhibited specific activity comparable to that observed by titration of purified PBAF (Fig. 2c, lanes 14–16). By contrast, resin from mock immunoprecipitates provided no cofactor activity (lanes 1, 2). Likewise, beads containing only affinity-purified anti-BAF180 antibody neither stimulated nor inhibited transcription (lanes 5–8). Taken together, these results indicate that a complex containing BAF180, defined as PBAF, is necessary and may be sufficient to serve as an important transcriptional cofactor in these reconstituted assays.

We next asked if other ATP-dependent chromatin-remodelling complexes might be able to satisfy the functional requirement exhibited by our *in vitro* transcription assay. Because the human SWI/SNF-A (SWI/SNF) and -B (PBAF) complexes are related^{3,7,12}, we wanted to determine whether SWI/SNF could replace PBAF as a transcriptional cofactor. These complexes share many subunits and could be expected to behave similarly during column chromatography, so it was unclear whether we could cleanly separate PBAF

from SWI/SNF. However, using western blot assays with antibodies directed against SWI/SNF- and PBAF-specific subunits (BAF250¹⁷ and BAF180² respectively), we were able to devise a purification scheme (Fig. 3a) yielding ~6,000-fold purification of SWI/SNF from human nuclear extracts (see Supplementary Information). As expected, the presence of BAF250 in SWI/SNF and its absence in PBAF was clearly evident when we directly compared our most purified fractions using western blot and SDS-PAGE silver-staining analysis (Fig. 3b, c, lane 3 asterisk). By contrast, BAF180 was detected only in PBAF and not SWI/SNF preparations (Fig. 3b, c, lane 1, black triangle). Thus, our experiments confirm the specific association of these subunits with distinct complexes.

Many chromatin-remodelling complexes have been described and some have been proposed to influence transcription from chromatin templates, so we sought to include other complexes in a comparative analysis. The imitation switch (ISWI) ATPase¹⁸ and its orthologous polypeptides represent another family of ATP-dependent chromatin-remodelling complexes including NURF¹⁸, ACF⁴ and RSF¹⁹. ACF was recently reconstituted from recombinant ISWI and Acl1 polypeptides, which were both found to be necessary and sufficient for ATP-dependent chromatin assembly and

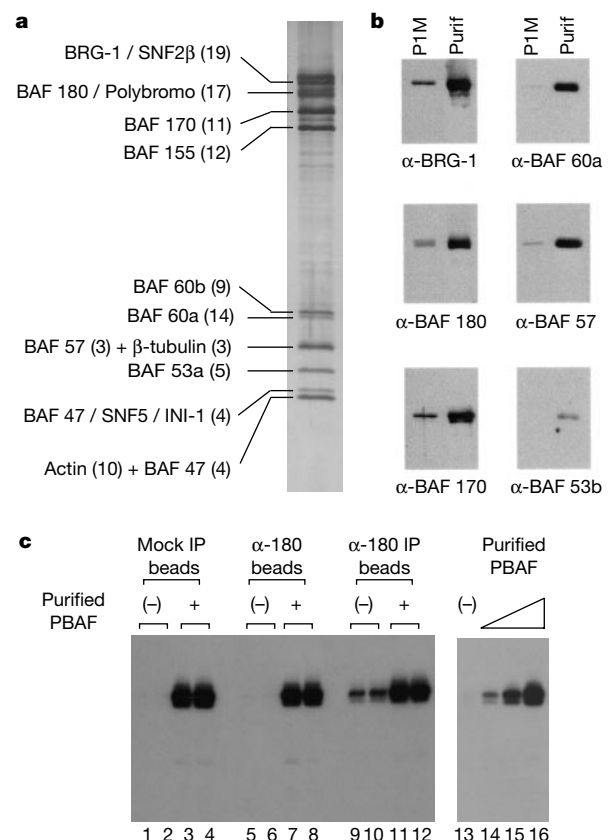


Figure 2 Chromatin-dependent transcription cofactor defined as PBAF. **a**, Mass spectrometric identification of polypeptides. Poros-HS fraction analysed as in Fig. 1e with gene products denoted and number of matching tryptic peptides in parentheses. **b**, Multiple BAFs specifically detected in the highly purified cofactor. 3 μ g phosphocellulose 1 M KCl fraction (P1M) and 0.15 μ g purified PBAF (Purif) separated by 7% SDS-PAGE and probed with antibodies denoted. **c**, Anti-BAF180 immunopurified complexes promote transcription. Beads from mock immunoprecipitation (lanes 1–4), beads containing affinity-purified anti-BAF180 antibody alone (lanes 5–8) or same used in immunoprecipitation of PBAF (lanes 9–12) assayed by *in vitro* transcription as in Fig. 1c, d together in the absence of PBAF (lanes 1, 2, 5, 6, 9, 10) or supplemented with purified PBAF (~1.5 nM, lanes 3, 4, 7, 8, 11, 12). PBAF titration (~0.1, 0.3 and 1 nM) in the absence of beads is shown (lanes 14–16).

remodelling⁴. We therefore chose to compare the activity of our PBAF and SWI/SNF complexes with purified recombinant ACF (Fig. 3c, lane 5).

First, we compared the specific ATP-dependent chromatin-remodelling activities of these three highly purified cofactors. To enable a direct comparison of chromatin remodelling under conditions relevant to our *in vitro* transcription assays, we tested these activities in the context of our nuclear receptor-dependent chromatin templates. We used a standard micrococcal nuclease pattern disruption assay to detect nucleosome mobilization at purified *in vitro* assembled chromatin templates upon hybridization with a probe containing the promoter sequence. Several experiments involving titration of each complex over a 30-fold range showed that the specific ATP-dependent chromatin-remodelling activities

of PBAF (Fig. 4a, lanes 5–8) and SWI/SNF (lanes 9–12) preparations varied by no more than 1.5-fold. Likewise, the specific activity of recombinant ACF was assessed to be within twice that of PBAF and SWI/SNF (lanes 13–16). We note that chromatin remodelling by these three different complexes was found to be neither nuclear receptor dependent nor selectively localized to the promoter region

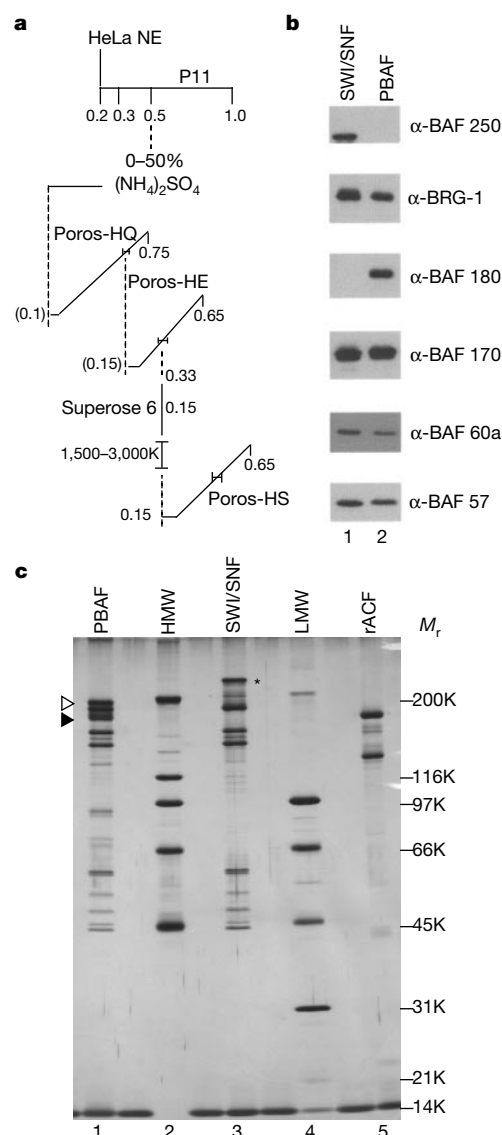


Figure 3 Purification of other chromatin-remodelling complexes. **a**, Chromatography scheme for purification of SWI/SNF. Numbers indicate KCl molarity, parentheses denote dialysis or dilution, and brackets designate pools. **b**, Comparative western blot analysis of SWI/SNF (2 μ l, ~0.45 μ g, lane 1) and PBAF (3 μ l, ~0.45 μ g, lane 2) as in Fig. 2b. **c**, Three highly purified chromatin-remodelling complexes. Silver-stained 5–15% gradient SDS-PAGE of PBAF (4 μ l, ~0.6 μ g, lane 1), high molecular weight standards (HMW, 0.6 μ g, Biorad, lane 2), SWI/SNF (3 μ l, ~0.7 μ g, lane 3), low molecular weight standards (LMW, 0.6 μ g, Biorad, lane 4), and recombinant ACF (1 μ l, ~0.25 μ g, lane 5). PBAF-specific polypeptides BAF200 (white triangle) and BAF180 (black triangle) as well as the SWI/SNF-specific BAF250 polypeptide (asterisk) are denoted.

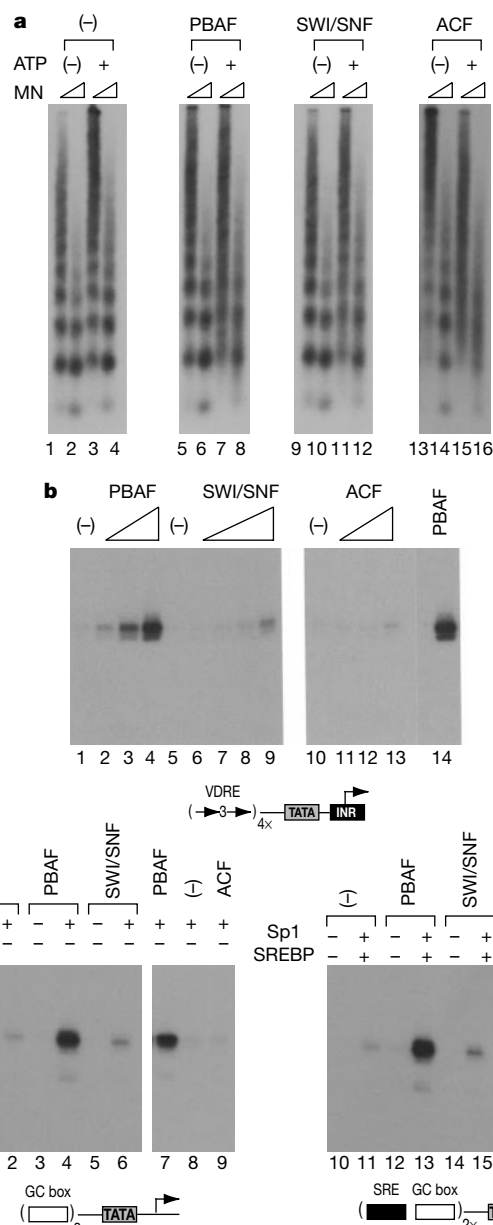


Figure 4 Specific requirement for PBAF chromatin cofactor by human enhancer binding factors. **a**, Chromatin remodelling by three highly purified complexes. Micrococcal nuclease (MN) pattern disruption of purified VDRE-containing chromatin templates by 1 nM PBAF (lanes 5–8), 1 nM SWI/SNF (lanes 9–12), or 3 nM rACF (lanes 13–16) incubated with 20 nM RXR + VDR and 100 nM 1,25(OH)₂D₃ in the absence or presence of ATP (top) detected with a promoter sequence probe as described in Methods. **b**, Selective utilization of PBAF in nuclear receptor-directed transcription. Reconstituted *in vitro* transcription reactions as in Figs 1c, d, supplemented with PBAF (0.1 nM, 0.3 nM, 1 nM, lanes 2–4, 14), SWI/SNF (0.1 nM, 0.3 nM, 1 nM, 3 nM, lanes 6–9), or ACF (1 nM, 3 nM, 10 nM, lanes 11–13). **c**, PBAF-specificity for transcription directed by two other human activators. Reconstituted *in vitro* transcription reactions with Sp1 and SREBP-1a as indicated (top) together with PBAF (1 nM, lanes 3, 4, 7, 12, 13, 16), SWI/SNF (1 nM, 5, 6, 14, 15), or ACF (3 nM, lanes 9, 18) programmed with pre-assembled chromatin templates (bottom).

as comparable disruption was detected by distal probes up to 2 kilobases (kb) from the promoter (data not shown).

Surprisingly, when these highly purified and active chromatin-remodelling complexes were used in our nuclear receptor-dependent *in vitro* transcription assay, only PBAF efficiently supported activated transcription (Fig. 4b, lanes 1–9); SWI/SNF was largely inactive in providing cofactor function. With several titration experiments normalized for relative chromatin-remodelling activity, we estimate that the specific transcriptional cofactor activity of PBAF is 20 to 30-fold higher than SWI/SNF. We suspect that the very low levels of cofactor activity detected for high concentrations of SWI/SNF (~3 nM, Fig. 4b, lane 9) may be due to trace contamination of PBAF in our SWI/SNF preparations which could not be detected by western blot. Adding purified SWI/SNF to transcription reactions directly supplemented with PBAF neither stimulated activity nor inhibited transcriptional activation (data not shown). These experiments suggest that SWI/SNF and PBAF are functionally different and exhibit a degree of specificity not previously observed. Further support for differential specificity of chromatin-remodelling complexes in transcription was obtained by comparing PBAF and ACF (Fig. 4b, lanes 10–14). Even at high concentrations of ACF (10 nM, ~1:1 per nucleosome), we observed minimal transcription stimulation (2-fold, lane 13). From several experiments, we estimate PBAF has transcriptional cofactor activity about 100-fold greater than that of ACF when normalized for chromatin remodelling. Notably, the highest ACF concentrations tested inhibited PBAF-dependent transcription by 50–70% (data not shown).

We had thus far used only one class of transcription activators (nuclear receptors), so it seemed prudent to test others. We examined transcription from two additional templates: one with cognate binding sites for Sp1 and a second that is synergistically responsive to Sp1 and SREBP-1a²⁰. Transcription directed from these templates was previously found to require the structurally related coactivator complexes CRSP⁶ and ARC⁵. In the presence of non-limiting amounts of these cofactors, activated transcription from both templates exhibited a strong dependence on PBAF (15 to

30-fold, Fig. 4c, lanes 4, 7, 13, 16). This cofactor requirement could not be complemented by SWI/SNF (~2-fold, lanes 6, 15), nor by ACF (no effect, lanes 9, 18). Taken together, these results establish that ATP-dependent chromatin-remodelling cofactor complexes are not functionally interchangeable, and suggest that the transcriptional role of PBAF may not be limited to its chromatin-remodelling activity.

We next investigated the relative requirements for other cofactors in our assay. It has been suggested that the TAFs in TFIID can be dispensable for nuclear receptor-directed transcription under certain conditions²¹. To assess the relative requirements for TAFs and BAFs, we examined transcription mediated by nuclear receptors on both chromatin (Fig. 5a, lanes 1–12) and naked DNA templates (lanes 13–24) each in the presence of affinity-purified ARC and a CRSP-containing fraction. Transcription from a chromatin template required both PBAF and TFIID (Fig. 5a, lanes 7–9), whereas purified TBP alone failed to substitute for TFIID (lanes 10–12). Predictably, nuclear receptor-activated transcription from a naked DNA template also required the TAF subunits of TFIID (Fig. 5a, lanes 13–15, 19–21). By contrast, PBAF was required for transcription only from chromatin templates (Fig. 5a, lanes 7–9) and not from naked DNA (lanes 13–15). Importantly, ligand-dependent activation by RXR + VDR required both a chromatin template and PBAF (Fig. 5a, lanes 8, 9). These findings suggest that TAFs and PBAF are both required for ligand-dependent transcription elicited by nuclear receptors, but each probably provides a distinct function.

It has been suggested that different classes of cofactors may be functionally redundant for transcription^{22,23}. Given the strong dependence on PBAF observed in our transcription assays, we wanted to examine its potential to replace the coactivators CRSP⁶ and ARC⁵/DRIP⁸. These coactivators were previously identified as important for transcription activation by the gene regulators used in this study. Using nuclear receptor-directed chromatin templates, experiments were performed in the absence (Fig. 5b, lanes 1, 2, 7, 8) or presence (lanes 3–6) of affinity-purified ARC and a CRSP-containing fraction (hereafter collectively designated ARC). These experiments revealed that both PBAF and ARC are required for ligand-dependent activated transcription by nuclear receptors from chromatin templates (lanes 5, 6). Similar results were obtained using Sp1- and SREBP-dependent templates (see Supplementary Information). Thus, our findings suggest that at least three distinct cofactor complexes, PBAF, ARC and TFIID, are interdependent for activated transcription by several human enhancer binding factors at chromatin templates.

Despite the identification of many ATP-dependent chromatin-remodelling and other chromatin-modifying activities from yeast to human, most have only been indirectly linked to transcriptional control *in vivo*^{9,24,25} or have exhibited modest effects on transcription *in vitro*^{26,27}. The unbiased *de novo* purification of PBAF, but not structurally related complexes^{28,29}, using an integrated chromatin-dependent transcription assay suggests that there may be more specificity to the function of apparently similar cofactors than previously anticipated. Transcription studies *in vivo*^{24,30} have largely been unable to distinguish functional specificity between highly similar complexes. It was therefore intriguing to find that other ATP-dependent chromatin-remodelling complexes, such as SWI/SNF and ACF, were unable to complement PBAF transcriptional cofactor activity. These findings suggest that different chromatin-remodelling complexes may perform quite specific functions that are not interchangeable in supporting transcription. Because chromatin remodelling was found to be both activator- and promoter-independent, it remains unclear under what circumstances these chromatin-remodelling complexes might be recruited to the template. Thus, it will be interesting to determine when gene-specific regulators and their attendant cofactors are required in a given transcription process¹. Whether some chromatin-remodelling activities are required strictly to facilitate template accessibility, to

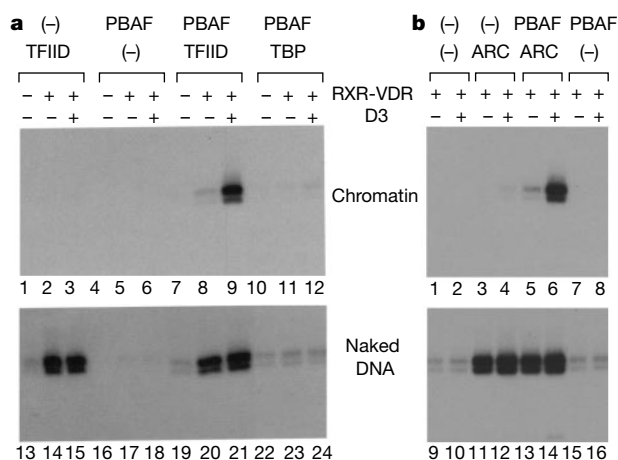


Figure 5 BAF, TAF, and ARC requirement for ligand-dependent activated transcription. **a**, Integral TFIID dependence and template-selective PBAF requirement. Reconstituted *in vitro* transcription reactions lacking TFIID supplemented with ARC + CRSP (Methods), RXR + VDR and ligand as indicated (top), 1 nM PBAF (lanes 4–12 and 16–24) and 0.6 nM TFIID (lanes 1–3, 7–9, 13–15 and 19–21) or TBP (lanes 10–12 and 22–24) programmed with pre-assembled chromatin (lanes 1–12) or naked DNA (lanes 13–24) templates. **b**, ARC and PBAF required for chromatin-dependent activated transcription. Reconstituted *in vitro* transcription reactions containing TFIID supplemented with RXR + VDR and ligand as indicated (top) together with PBAF (1 nM, lanes 5–8, 13–16) and ARC + CRSP (ARC, lanes 3–6, 11–14) programmed with pre-assembled chromatin (lanes 1–8) or naked DNA templates (lanes 9–16).

induce specialized template conformations, or to assemble transcriptionally active complexes (perhaps by mediating interactions with other cofactors) remain open questions. Well defined transcription and remodelling assays that integrate various chromatin-associated functions should provide a means to discriminate related cofactor complexes and help to determine their biochemical roles. □

Methods

Transcription factors, templates and *in vitro* transcription

The reconstituted general transcription system includes: recombinant bacterially expressed TFIIA, -B, -E, and -F, purified human RNA Pol II prepared as described²⁰, and immunopurified human TFIID and TFIIF prepared as described⁴. A CRSP-containing fraction was prepared as described²⁰. ARC and CBP were affinity-purified from nuclear extracts as described⁷ and further purified by Poros 20 HE-1 (PerSeptive Biosystems) heparin chromatography and glycerol gradient sedimentation for use in some experiments. VDR and Flag-RXR were prepared as described⁸, Flag-PPAR γ was expressed in bacteria and purified similarly except buffer D (25 mM HEPES of pH 7.9, 0.2 mM EDTA, 2 mM MgCl₂, 10% glycerol + KCl) was used and eluted in 0.15 M KCl. Purification of SREBP-1a and Sp1 were described²⁰. The cofactor fraction in which PBAF activity was initially detected (Cof, Fig. 1a) was obtained from HeLa nuclear extracts followed by P11-phosphocellulose chromatography (Whatman, 15 mg ml⁻¹ resin) eluted at 1 M KCl, subsequently dialysed to 0.13 M KCl, passed through the weak anion exchanger DE-52 (Whatman, 3 mg ml⁻¹ resin), applied to Poros 20 HE-1, and developed with a linear gradient from 0.13 M to 1 M KCl. Active fractions eluting at 0.37–0.46 M KCl were pooled and dialysed to 0.1 M KCl. A detailed description of the purification of PBAF and SWI/SNF is presented in the Supplementary Information. Recombinant ACF was expressed and purified as described⁴. The G3B-CAT, LDLR-CAT, and VDRE $\times$ 4-B/INR-CAT templates were described^{16,20}. The PPRE $\times$ 2-B/INR-CAT template was generated by replacing vitamin D3 responsive element (VDRE) sequences in VDRE $\times$ 4-B/INR-CAT with oligonucleotides containing a peroxisome proliferator responsive element (PPRE) from the rat *acyl CoA oxidase* gene. Each template was assembled into chromatin or used as mock assembled naked DNA templates as described⁸. *In vitro* transcription reactions contained: 0.5 nM DNA templates, 0–5 nM Sp1 and SREBP-1a or 0–20 nM nuclear receptors and ligands as indicated in the figure legends, 40 nM rIIA, 5 nM rIIB, 20 nM rIIE, 20 nM rIIF, 2 nM RNA Pol II, 0.1 nM TFIIF, and 0.6 nM TFIID or rTBP in a 25- μ l volume. Non-limiting amounts of coactivators (0.5 nM ARC and 6 μ g ml⁻¹ of a CRSP-containing fraction) were used as indicated. Reactions were performed as described⁸ except activators were added after 5.5 h of chromatin assembly, pre-incubated 10 min before addition of general factors, coactivators, and chromatography fractions, further incubated at 28.5 °C 15 min before NTPs (0.5 mM final) and incubated 30 min before RNA isolation and primer extension.

Polypeptide identification and immunomethods

Peak Poros-HS fractions were pooled (~10 μ g protein), concentrated by TCA precipitation, separated by SDS–PAGE, stained, excised, digested with trypsin, and extracted. Peptide pools from each gel slice were analysed by matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS; Bruker Reflex III). Selected mass values were used to search protein databases linked to PROWL (Rockefeller University) using ProFound (http://129.85.19.192/profound_bin/WebProFound.exe) and protein databases linked to ExPASy (Swiss Institute of Bioinformatics, Geneva) using PeptIdent (<http://www.expasy.ch/tools/peptident.html>). Experimental masses shared by no more than two bands corresponding to zero or one missed theoretical cleavages without modifications within 0.5 atomic mass units (typically 0.2 a.m.u.) were included in Fig. 2a. Antibodies were used at 1:1,000 for western blot analysis by standard methods and visualized by enhanced chemiluminescence (Amersham-Pharmacia). A detailed description of immunoprecipitation for the transcription on beads experiment (Fig. 2c) is presented in the Supplementary Information.

Chromatin-remodelling assays

Chromatin templates were prepared exactly as for transcription assays, incubated with 0.05% Sarkosyl for 10 min at 22 °C and purified by CL4B gel filtration chromatography. Purified templates (~0.5 nM) and chromatin-remodelling factors (0–10 nM) were incubated together in the absence or presence of 2 mM ATP in 62.5 mM KCl, 12.5 mM HEPES (pH 7.7), 8% glycerol, 6.5 mM MgCl₂, 0.1 mM EDTA, 0.5% each polyvinyl alcohol (average *M*_w 9K–10K) and polyethylene glycol (15K–20K), 0.005% NP-40, plus 10 μ g ml⁻¹ each BSA and lysozyme carrier protein in a reaction volume of 75 μ l for 1 h at 28.5 °C. Remodelled templates were subjected to digestion with micrococcal nuclease (0.12–0.6 U ml⁻¹) for 10 min at 22 °C, deproteinized, and purified. Approximately 20 ng of DNA per reaction was separated on 1.2% agarose gels, transferred to nylon (Hybond-N+, Amersham-Pharmacia), and hybridized with oligonucleotide probes.

Received 30 August; accepted 18 October 2001.

- Lemon, B. & Tjian, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14**, 2551–2569 (2000).
- Xue, Y. T. *et al.* The human SWI/SNF-B chromatin-remodelling complex is related to yeast Rsc and localizes at kinetochores of mitotic chromosomes. *Proc. Natl Acad. Sci. USA* **97**, 13015–13020 (2000).
- Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* **370**, 477–481 (1994).

- Ito, T. *et al.* ACF consists of two subunits, Acl1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* **13**, 1529–1539 (1999).
- Näär, A. M. *et al.* Composite coactivator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**, 828–832 (1999).
- Ryu, S., Zhou, S., Ladurner, A. G. & Tjian, R. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**, 446–450 (1999).
- Kingston, R. E. & Narlikar, G. J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352 (1999).
- Rachez, C. *et al.* Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* **398**, 824–828 (1999).
- Kwok, R. P. *et al.* Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature* **370**, 223–226 (1994).
- Ge, H. & Roeder, R. G. Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78**, 513–523 (1994).
- Merino, A., Madden, K. R., Lane, W. S., Champoux, J. J. & Reinberg, D. DNA topoisomerase I is involved in both repression and activation of transcription. *Nature* **365**, 227–232 (1993).
- Wang, W. *et al.* Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev.* **10**, 2117–2130 (1996).
- Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170–174 (1993).
- Muchardt, C. & Yaniv, M. A human homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and *Drosophila* brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* **12**, 4279–4290 (1993).
- Reyes, J. C., Muchardt, C. & Yaniv, M. Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. *J. Cell Biol.* **137**, 263–274 (1997).
- Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422–1425 (2000).
- Nie, Z. *et al.* A specificity and targeting subunit of the human SWI/SNF family-related chromatin-remodeling complex. *Mol. Cell Biol.* **20**, 8879–8888 (2000).
- Tsukiyama, T., Daniel, C., Tamkun, J. & Wu, C. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* **83**, 1021–1026 (1995).
- LeRoy, G., Orphanides, G., Lane, W. S. & Reinberg, D. Requirement of RSF and FACT for transcription of chromatin templates *in vitro*. *Science* **282**, 1900–1904 (1998).
- Näär, A. M. *et al.* Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a *in vitro*. *Genes Dev.* **12**, 3020–3031 (1998).
- Fondell, J. D., Guermah, M., Malik, S. & Roeder, R. G. Thyroid hormone receptor-associated proteins and general positive cofactors mediate thyroid hormone receptor function in the absence of the TATA box-binding protein-associated factors of TFIID. *Proc. Natl Acad. Sci. USA* **96**, 1959–1964 (1999).
- Gu, W. *et al.* A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* **3**, 97–108 (1999).
- Lee, T. I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405**, 701–704 (2000).
- Fryer, C. J. & Archer, T. K. Chromatin remodelling by the glucocorticoid receptor requires the BRG1 complex. *Nature* **393**, 88–91 (1998).
- Krebs, J. E., Kuo, M. H., Allis, C. D. & Peterson, C. L. Cell cycle-regulated histone acetylation required for expression of the yeast HO gene. *Genes Dev.* **13**, 1412–1421 (1999).
- Armstrong, J. A., Bieker, J. J. & Emerson, B. M. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF *in vitro*. *Cell* **95**, 93–104 (1998).
- Kraus, W. L. & Kadonaga, J. T. p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation. *Genes Dev.* **12**, 331–342 (1998).
- Kal, A. J., Mahmoudi, T., Zak, N. B. & Verrijzer, C. P. The *Drosophila* brahma complex is an essential coactivator for the trithorax group protein zeste. *Genes Dev.* **14**, 1058–1071 (2000).
- Sif, S., Saurin, A. J., Imbalzano, A. N. & Kingston, R. E. Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev.* **15**, 603–618 (2001).
- Bultman, S. *et al.* A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* **6**, 1287–1295 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Bell and M. Haggart for technical assistance; W. Wang for providing affinity-purified antibodies raised against BAF180; G. Crabtree for antibodies raised against BRG1 and multiple BAFs; N. Tanese for antibodies raised against BAF250; R. Kingston and A. Imbalzano for antibodies raised against BRM; D. Fyodorov and J. Kadonaga for providing recombinant viruses for expression of ACF and for purified rACF used in initial experiments; S. Ryu for purified TFIIF, rPC4, rTopo I, and rTBP, as well as the G3B-CAT plasmid; A. Näär for purified rSREBP-1a as well as LDLR-CAT and GST-activation domain expression plasmids; J.-L. Chen for the PPAR γ cDNA and ligand; L. Freedman, R. Freiman, J. Kadonaga, P. Kaufman, M. Marr, D. Taatjes, and W. Wang for comments on the manuscript, and all members of our laboratory for helpful discussions and technical advice. B.L. was recipient of a National Research Service Award during most of this work.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.T. (e-mail: jmlim@uclink4.berkeley.edu).

Stimulatory effect of splicing factors on transcriptional elongation

Yick W. Fong & Qiang Zhou

Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720-3206, USA

Transcription and pre-mRNA splicing are tightly coupled gene expression events in eukaryotic cells^{1,2}. An interaction between the carboxy-terminal domain of the largest subunit of RNA polymerase (Pol) II and components of the splicing machinery is postulated to mediate this coupling³⁻⁵. Here, we show that splicing factors function directly to promote transcriptional elongation, demonstrating that transcription is more intimately coupled to splicing than previously thought. The spliceosomal U small nuclear ribonucleoproteins (snRNPs) interact with human transcription elongation factor TAT-SF1 (refs 6-9) and strongly stimulate polymerase elongation when directed to an intron-free human immunodeficiency virus-1 (HIV-1) template. This effect is likely to be mediated through the binding of TAT-SF1 to elongation factor P-TEFb¹⁰, a proposed component of the transcription elongation complex^{11,12}. Inclusion of splicing signals in the nascent transcript further stimulates transcription, supporting the notion that the recruitment of U snRNPs near the elongating polymerase is important for transcription. Because the TAT-SF1-U snRNP complex also stimulates splicing *in vitro*, it may serve as a dual-function factor to couple transcription and splicing and to facilitate their reciprocal activation.

The Pol II C-terminal domain (CTD) is hyperphosphorylated during the transcription cycle at a time coincident with processive polymerase elongation¹³. Phosphorylation of the CTD during elongation is carried out primarily by P-TEFb, a heterodimer of CDK9 and cyclin T1 (CYCT1)¹⁰. P-TEFb is also a cellular cofactor for the HIV-1 TAT protein, and the TAT-P-TEFb complex stimulates HIV-1 transcriptional elongation by interacting with the TAR RNA structure located at the 5' end of the nascent viral transcript^{14,15}. Although the cyclin box located in the amino-terminal half of CYCT1 is essential as it contacts CDK9, TAT and TAR¹⁶, the CYCT1 C-terminal half has also been shown to contribute significantly to both basal and TAT-stimulated HIV-1 transcription^{17,18}.

The importance of the CYCT1 C-terminal domain prompted us to identify and analyse transcription factors that may associate with this domain. We therefore incubated nuclear extract of HeLa cells with immobilized GST or GST-CYCT1-C, which contained a CYCT1 C-terminal fragment (amino acids 402-701; ref. 18). Compared with the GST-depleted extract, extract depleted with GST-CYCT1-C showed a significant decrease (about ninefold on average) in basal as well as TAT-dependent HIV-1 transcription from both templates: pHIV+TAR-G400, which contained the wild-type TAR element; and pHIVΔTAR-G100, with a mutant TAR¹⁹ (Fig. 1a). Because the level of TAT activation (about eightfold) was largely unaffected by the depletion, CYCT1-C probably removed general transcription activity.

In nuclear extract of HeLa cells, CYCT1-C has been shown to interact with Pol II and TAT-SF1¹⁸, one or both of which may contribute to the activity depleted by CYCT1-C. TAT-SF1 has been identified as a TAT cofactor as well as a general transcription elongation factor⁶⁻⁹. Because it is highly enriched in a partially purified Q-Sepharose fraction of HeLa nuclear extract that contains a small portion of HeLa nuclear proteins (3-4%), including Pol II⁸, we examined whether this fraction can complement the CYCT1-C-depleted extract in transcription. HIV-1 transcription was restored by the addition of the Q fraction pre-depleted with GST, but not with GST-CYCT1-C (Fig. 1a), suggesting that the activity depleted

from HeLa nuclear extract by CYCT1-C also exists in the Q fraction. Western analysis revealed a quantitative removal of TAT-SF1 from the CYCT1-C-depleted Q fraction (Fig. 1b). However, CYCT1-C only partially removed SPT5, an elongation factor reported to bind TAT-SF1 (ref. 6), and it removed very little Pol II or the TFIIF subunit RAP30. Thus, Pol II was probably not responsible for the activity depleted by CYCT1-C.

Rather, the CYCT1-C-depleted activity seemed to associate with TAT-SF1, because the immunoprecipitation of TAT-SF1 and its associated factors from the Q fraction with anti-TAT-SF1, but not with preimmune antibodies, restored both basal and TAT-dependent HIV-1 transcription to a CYCT1-C-depleted nuclear extract (Fig. 1c). Because these reactions measured transcription from two G-less cassettes (G-400 and G-100) located about 1 kilobase (kb) downstream of the HIV-1 promoter¹⁹, the anti-TAT-SF1 immunoprecipitates were postulated to transactivate at the level of elongation. To confirm this, we performed an assay involving discontinuous

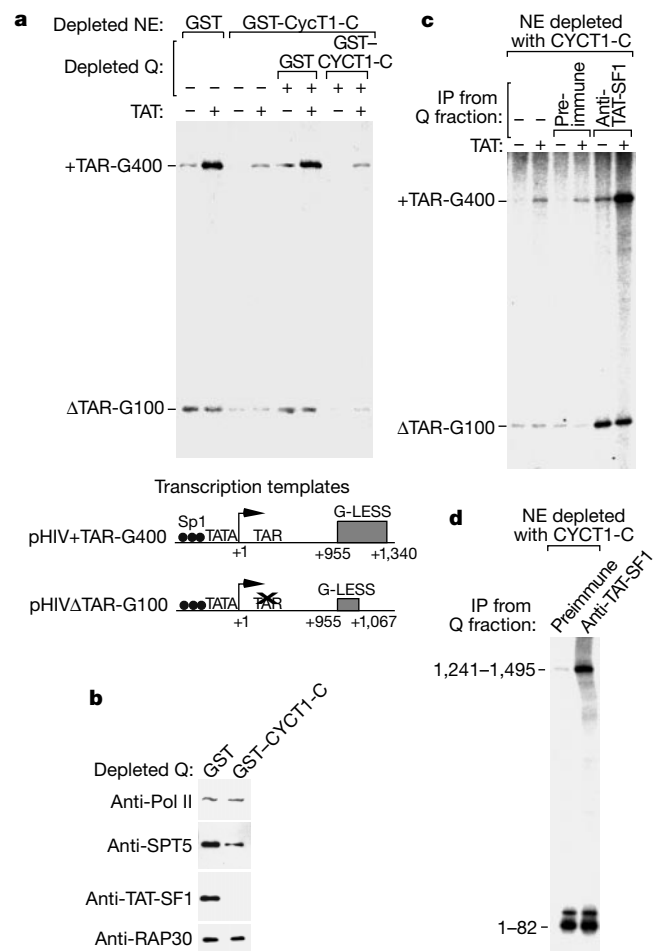


Figure 1 The CYCT1 C-terminal domain interacts with a TAT-SF1-associated transcription elongation activity. **a**, HeLa nuclear extract (NE) and the Q-Sepharose fraction (Q) were depleted with immobilized GST or GST-CYCT1-C and then incubated with templates pHIV+TAR-G400 and pHIVΔTAR-G100 in reactions with or without TAT. The RNase T1-resistant G-less transcripts derived from the two templates are indicated. **b**, The levels of the indicated proteins in Q depleted with GST or GST-CYCT1-C were analysed by western blotting. **c**, The Q fraction was subjected to immunoprecipitation with preimmune or anti-TAT-SF1 antibody beads. The immunoprecipitates (IP) were analysed in reactions containing the CYCT1-C-depleted NE as in **a**. **d**, A discontinuous hybridization followed by RNase protection assay²⁰ was performed to analyse the level of transcription from template pHIV+TAR at two different distances from the initiation site. Reactions contained the CYCT1-C-depleted NE and the indicated immunoprecipitates. Numbers denote the 5' and 3' extent of the protected RNA fragments.

hybridization followed by RNase protection²⁰ to measure transcription at two different locations from the initiation site (Fig. 1d). Addition of the TAT-SF1 immunoprecipitates to the CYCT1-C-depleted extract only weakly stimulated the promoter-proximal transcription (nucleotides 1–82), but enhanced the promoter-distal transcription (nucleotides 1,241–1,495) by 20-fold, revealing a predominant effect on elongation.

Sequence comparisons indicate that of all vertebrate proteins, human and mouse TAT-SF1 proteins are most homologous to CUS2 of *Saccharomyces cerevisiae*. The homology lies within the two atypical RNA recognition motifs (RRMs) and their flanking regions located in the N-terminal half of TAT-SF1 (Fig. 2a). The extensive acidic C-terminal domain in human TAT-SF1 is reduced to a 25-amino-acid segment in CUS2. CUS2 was identified as a suppressor of a U2 snRNA mutation and shown to regulate the spliceosomal formation^{21,22}. In splicing extracts, CUS2 associates with U2 snRNA and splicing factor SF3a²¹.

The homology between TAT-SF1 and CUS2 suggests that TAT-SF1 may also bind splicing factors, which could be responsible for the observed stimulatory effect on transcriptional elongation. To test these hypotheses, a TAT-SF1 mutant containing a point mutation (Y136D) in the first RRM was generated (Fig. 2a). A corresponding CUS2 mutation (Y48D) destroys the CUS2–U2 snRNA interaction and abrogates CUS2 activity²¹. Because the extended C-terminal half of TAT-SF1 is absent in CUS2, a truncated TAT-SF1, termed Δ C (amino acids 1–380; Fig. 2a) was also generated. Flag-tagged wild-type and mutant TAT-SF1 and their associated proteins (TAT-SF1–Flag immunoprecipitates) were affinity purified from the nuclear extracts of transfected 293T cells and examined by western blotting. Consistent with CUS2 being a splicing factor, wild-type TAT-SF1–Flag interacted with the U1 snRNP-specific protein U1 70K, U2-specific protein U2B'', and the common Sm proteins B and B' (Fig. 2b). The Y136D mutation destroyed all of these interactions, whereas the Δ C deletion showed no effect. Northern hybridization (Fig. 2c) further demonstrated that both the TAT-SF1 immunoprecipitates isolated from the Q fraction and the transfected wild-type TAT-SF1–Flag purified from the nuclear extract associated with all five spliceosomal U snRNAs. Again, the Y136D mutation disrupted these interactions. Compared with wild-type TAT-SF1–Flag, Δ C interacted with less U5 and U4/U6. Thus, the TAT-SF1 RRM domain seemed to be primarily responsible for the interactions of TAT-SF1 with the snRNPs, although the C-terminal acidic domain also contributed to some of the interactions.

In reactions containing the CYCT1-C-depleted HeLa nuclear extract, addition of the wild-type TAT-SF1–Flag immunoprecipitate increased HIV-1 transcription (Fig. 2d). In contrast, transcription was slightly decreased by the Δ C preparation and significantly inhibited by the Y136D preparation, suggesting that both the RRM and the acidic C-terminal domains of TAT-SF1 are required for transcription. To explain these differences, the wild-type and mutant TAT-SF1–Flag immunoprecipitates were tested for interacting with immobilized GST–CYCT1-C (Fig. 2e). Whereas both the wild-type and Y136D preparations showed active binding, the Δ C complex failed, revealing a requirement for the TAT-SF1 C terminus in contacting P-TEFb. P-TEFb was reported to associate with the Pol II elongation complex along the HIV-1 template through position +529 (refs 11, 12), which may recruit the TAT-SF1–snRNP complex to Pol II through the CYCT1–TAT-SF1 interaction. Therefore, the lack of transactivation by Δ C could be attributed to its failure to associate with the elongating Pol II through CYCT1, and/or its decreased binding to U5 and U4/U6 snRNPs. Although Y136D did not associate with any snRNP, it interacted with CYCT1-C even stronger than did the wild-type protein (Fig. 2e). This may prevent the recruitment of a functional TAT-SF1–snRNP complex to Pol II and thus inhibit transcription.

These interpretations were based on the assumption that the TAT-SF1-associated splicing factors are capable of transactivation. To confirm this, the immobilized TAT-SF1 immunoprecipitates were treated with micrococcal nuclease and washed extensively to remove the disrupted snRNPs. Like Y136D, the material treated with micrococcal nuclease inhibited transcription mediated by the CYCT1-C-depleted extract (Fig. 3a), suggesting that TAT-SF1 alone is incapable of transactivation.

Further proof of the transactivating activity of the TAT-SF1–snRNP complex came from two depletion studies. First, we depleted HeLa nuclear extract with a monoclonal antibody (K121) recognizing the unique 2,2,7-trimethylguanosine (TMG) cap structure at the 5' end of all the spliceosomal U snRNAs except U6. This antibody has been used to affinity purify U snRNPs from HeLa extract^{23,24}. Compared with the control anti-HA-depleted extract, the anti-TMG-depleted extract showed a decrease in HIV-1 transcription (Fig. 3b), which was reversed by the addition of wild-type TAT-SF1–Flag immunoprecipitate. A reported direct interaction between yeast U2 snRNP and CUS2 (refs 21, 22) suggests

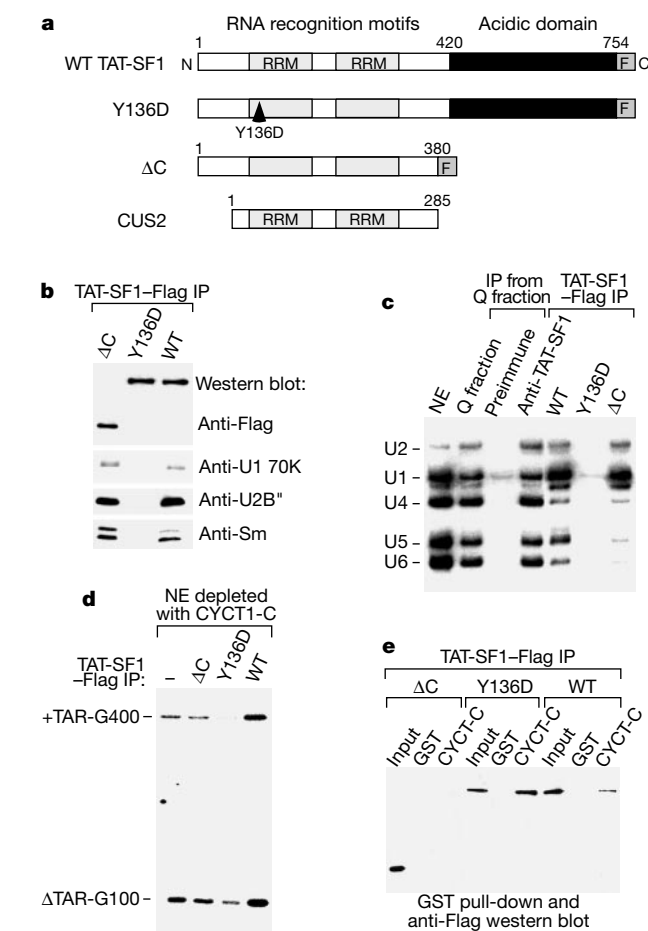


Figure 2 Both the N-terminal RRM and the C-terminal acidic domain of TAT-SF1 are required for transactivation. **a**, Domain structures of wild-type (WT) and mutant (Y136D and Δ C) TAT-SF1 and the yeast homologue of TAT-SF1, CUS2. **b**, Flag-tagged wild-type and mutant TAT-SF1 were affinity purified from the nuclear extracts of transfected 293T cells and analysed for their associated snRNP proteins by western blotting with the indicated antibodies. IP, immunoprecipitate. **c**, RNA recovered from the indicated sources was analysed by northern hybridization with snRNA-specific riboprobes. **d**, Affinity-purified wild-type and mutant TAT-SF1–Flag and their associated factors were analysed in transcription reactions containing the CYCT1-C-depleted HeLa nuclear extract (NE) as in Fig. 1a. Because TAT-SF1–snRNP affected general HIV-1 transcription, TAT was therefore omitted from this and the subsequent experiments, although the same two-template system was used for consistency. **e**, The indicated TAT-SF1–Flag immunoprecipitates were incubated with immobilized GST and GST–CYCT1-C. The bound proteins and 38% of the input materials were examined by anti-Flag western blotting.

The TAT-SF1-snRNP complex also complemented HeLa nuclear extract treated with micrococcal nuclease to splice a β -globin pre-mRNA (Fig. 3d). The treatment degraded all major RNA species in the extract, and abolished splicing²⁷. Addition of the TAT-SF1 immunoprecipitates isolated from the Q fraction restored splicing to the micrococcal-nuclease-treated extract. Thus, the TAT-SF1-snRNP complex seemed to function in both splicing and transcription.

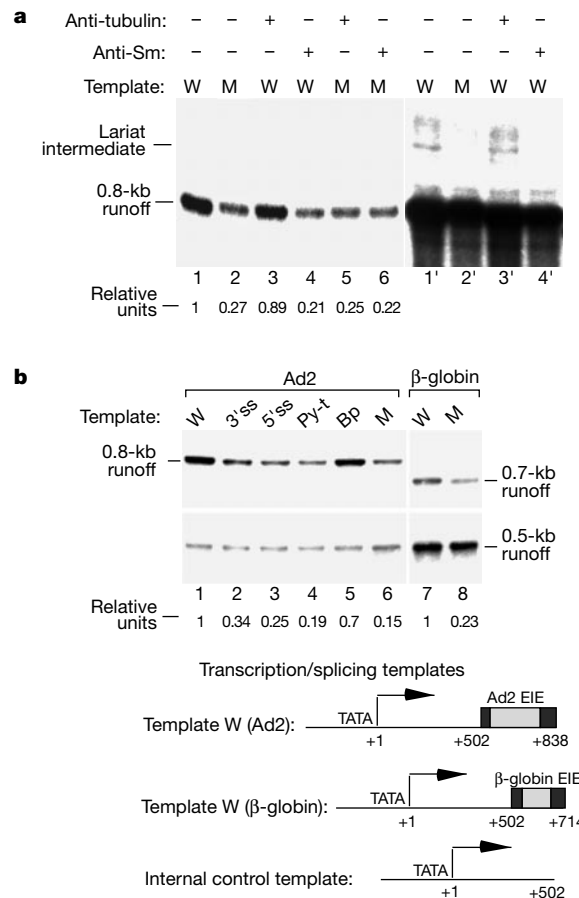


Figure 4 Splicing signals in nascent transcripts stimulate transcription. **a**, HeLa nuclear extract, linearized template W or M, and anti-tubulin or anti-Sm antibody were incubated in coupled transcription and splicing reactions, which produced ~0.8-kb runoff transcripts. Lanes 1'–4' were exposed eight times longer than lanes 1–4 to reveal the lariet intermediate. Template W and M contained, respectively, the wild-type and mutant (mutations at 5' and 3' splice sites and the polypyrimidine tract) Ad2 exon–intron–exon (EIE) sequences inserted downstream of the HIV-1 promoter. **b**, All reactions contained HeLa nuclear extract and two templates. The internal control template produced an intron-free runoff transcript of ~0.5 kb. The second set of templates contained Ad2 or β -globin EIE and produced intron-containing transcripts of ~0.8 or ~0.7 kb. Four indicated templates carried individual mutations at the 5' (5'ss) or 3' splice site (3'ss), the polypyrimidine tract (Py-t), or the branch point (Bp) in Ad2 EIE. Template M contained multiple mutations as described in **a**. Relative levels of transcription were normalized against the internal controls.

We used intron-free templates in all transcription reactions described thus far. Recruitment of TAT-SF1–snRNP to these templates is probably mediated by the interaction of TAT-SF1 with CYCT1–P-TEFb, which was shown to travel with the elongating polymerase^{11,12}. For intron-containing templates, however, the presence of splicing signals in the nascent transcript may attract additional snRNPs near Pol II to further stimulate transcription. To test this hypothesis, we inserted a 336-base pair (bp) adenovirus DNA with wild-type exon–intron–exon sequences (pSPAd²⁸) downstream of the HIV-1 promoter (Fig. 4b). This template was linearized and a runoff transcript of about 0.8 kb was expected. Compared with a mutant template containing the mutated 5' and 3' splice sites and the polypyrimidine tract, the wild-type template displayed increased transcription (Fig. 4a, lanes 1 and 2). A longer exposure revealed a low level of splicing of only the wild-type transcript, as evidenced by the generation of a splicing intermediate (lariat intermediate) (lanes 1' and 2'). Inhibition of splicing by the anti-Sm antibody²⁹ reduced transcription from the wild-type but not the mutant template. Furthermore, individual mutations of the 5' and 3' splice sites and the polypyrimidine tract, but not the branch point adenosine, largely impaired the ability of the adenovirus intron to stimulate transcription (Fig. 4b). Confirming an intron-dependent stimulation process, a similar level of shorter (0.5-kb) transcripts was produced from an internal control template linearized before the exon–intron insert. Finally, an inserted β -globin intron also increased transcription, revealing the generality of the phenomenon (Fig. 4b). On the basis of these results, we propose that the transcriptionally active TAT-SF1–snRNP complex can be recruited to the elongation complex through the binding of TAT-SF1 to P-TEFb and the binding of snRNPs to the nascent splicing substrate. Notably, only the later process seemed to be sensitive to anti-Sm inhibition.

Although the major events in eukaryotic gene expression have traditionally been analysed individually, recent evidence indicates that they are highly coupled inside the cell^{1,2}. The couplings described thus far, however, have largely been unidirectional, involving one upstream event (for example, splicing) affecting one or more downstream events (for example, mRNA export and decay). Our observations that the TAT-SF1-associated splicing factors and the splicing signals in the nascent transcript promote transcriptional elongation suggest that the coupling between transcription and splicing can be mutually beneficial. As P-TEFb is travelling with the elongation complex^{11,12}, the CYCT1–TAT-SF1 interaction may recruit TAT-SF1–snRNP to the elongating polymerase, increase the local concentration of splicing factors, and facilitate the assembly of a spliceosome when splice sites emerge from the polymerase. This mechanism, combined with the association of splicing factors with Pol II CTD^{3–5}, may explain the high efficiency of co-transcriptional splicing *in vivo*. Although splicing may benefit from the coupled transcription, our data suggest that the two processes can exhibit reciprocal synergism, providing an efficient mechanism for their coordinated regulation in response to changing physiological conditions. A stimulatory effect on gene expression by the presence of introns in the transcription unit has been observed in both yeast and mice³⁰. Future experiments will shed light on the mechanism by which the TAT-SF1-associated splicing factors stimulate transcriptional elongation. □

Methods

Depletion of nuclear extracts

We incubated 0.5 mg of HeLa nuclear extracts in D buffer (20 mM HEPES buffer at pH 7.9, 10% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.05% NP40, 1 mM dithiothreitol and 0.5 mM phenylmethyl sulphonyl fluoride) twice with 1 mg of immobilized GST or GST–CYCT1–C for a total of 4 h at 4 °C to obtain the CYCT1–C-depleted extracts, or three times with anti-TMG monoclonal antibody coupled to protein G–Sepharose for a total of 3 h at 4 °C to obtain the anti-TMG-depleted extracts. Depletion of U2 snRNP was carried out as described²⁵ except that nuclear extracts in D buffer were used.

Affinity purification of TAT-SF1 complexes

Flag-tagged TAT-SF1 proteins were transiently expressed in 293T cells. Nuclear extracts were prepared 48 h after transfection and dialysed against D buffer. TAT-SF1 complexes were affinity purified from the dialysed extracts by incubating with anti-Flag antibody beads (Sigma), followed by extensive washes with D buffer and Flag peptide elution.

Affinity purification of U2 snRNP

U2 snRNP was affinity purified essentially as described²⁵. Briefly, the Q fraction in D buffer was incubated with 2 nmol ml^{–1} biotinylated 2'-O-methyl-RNA antisense oligonucleotide (5'-GGCCGAGAAGCGAUAU-biotin-3') with sequences complementary to the first 14 nucleotides of U2 snRNA. The reaction mix was incubated at room temperature for 30 min and then mixed for 1.5 h at 4 °C with pre-blocked streptavidin agarose beads (Sigma). After washing with D buffer, U2 snRNP was eluted by the addition of a 2.5-fold excess of a displacement DNA oligonucleotide (5'-ATCGCTTCTCGGCC-3') in D buffer containing 0.1% NP40 and 1 mg ml^{–1} BSA and then concentrated.

Coupled *in vitro* transcription and splicing

A 336-bp Ad2 fragment derived from pSPAd²⁸ containing a shortened intron and the first and second leader exons were amplified by PCR and inserted between the *Xba*I and *Bam*HI sites of the template pHIV+TAR-G400¹⁹. Similarly, a 212-bp β -globin exon–intron–exon fragment was subcloned into the *Xba*I and *Bam*HI sites of pHIV+TAR-G400. Various splice-site mutant fragments were generated by PCR using the sets of the following oligonucleotides: (1) 5'-CATGCTAGACTGCGAGGGCCAGCTG-3'; (2) 5'-CATGTCTAGACTGCGAGGGCCAGCTGTTGGTGCATGTGACTCCCTCTCAAAA-3'; (3) 5'-CTAGGATCCCGTTCGAGGCGACGGG-3'; (4) 5'-GTCCTTTTTCCTCAACTCTCGCGTTGAGGACAACT-3'; (5) 5'-AGTTTGTCTCAACCGCAGAGTTGGAA AAAAAGGGAC-3'; (6) 5'-GATGATGTCATCTATCAGGAGGAAAGAAAGGAC AGCTCGCGTTGAGGAC-3'; (7) 5'-GTCCTCAACCGCAGCTGTCTTTTCC TTTCTCTGATAAGTATGACATCATC-3'; (8) 5'-GGTTTCCTTCTACTGGTGCAT ACTTAGCTGTCCCTTTTTCCTCA-3'; (9) 5'-TGGAAAAAAGGGACAGTAAAG GTATGACCACTAGAAGGAAACC-3'; (10) 5'-CTTGTGTGTGTCATCTTTAGAGGA GGAAAGGAAAGGAACTCTCGCGTT-3'; (11) 5'-AAGTATGACAAACAAGG AAACCTGGACAAACAGCGCCCTAGACGTGC-3'; (12) 5'-CATGTCTAGAAGTT GTGGTGTAGGCGCCCTG-3'; (13) 5'-CATGTCTAGAAGTTGGTGTGAGGCC TGGGCTCAGCTACATCAAGTTTACAAGACAGG-3'; (14) 5'-CTAGGATTCGACAG ATCCCAAAGGACTC-3'; (15) 5'-TTCTGTTAGGGCACTTGTCTCTCGCTATT GTCTATTTTCCCACTTGTCTGCTGG-3'; (16) 5'-ACAAAGTGCCTAACAG AAACCCAA-3'. The following fragments were generated: Ad-w, oligonucleotides 1 and 3; Ad-3'ss, 1, 5, 3 and 4; Ad-5'ss, 2 and 3; Ad-Py-t, 1, 7, 3 and 6; Ad-Bp, 1, 9, 3 and 8; Ad-m, 2, 11, 3 and 10; β -globin-w, 12 and 14; and β -globin-m, 13, 16, 14 and 15. The transcription templates used in the assay were linearized with *Not*I and *Bam*HI or with *Not*I and *Xba*I to generate runoff transcripts of ~0.8 kb (Ad2), ~0.7 kb (β -globin) and ~0.5 kb (internal control). *In vitro* transcription condition was modified to allow more-efficient splicing. Briefly, the reaction was scaled up to 25 μ l, containing 2% PEG3000, 3.6 mM MgCl₂, 0.3 mM ATP, 50 μ M GTP, 50 μ M UTP, 6.25 μ M CTP, 10 μ Ci [α -³²P]CTP (800 Ci mmol^{–1}; 1 μ l), 25 ng linearized templates, and 30–50 μ g HeLa nuclear extract.

Received 14 August; accepted 17 October 2001.

- Bentley, D. Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* **11**, 347–351 (1999).
- Hirose, Y. & Manley, J. L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429 (2000).
- McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
- Fong, N. & Bentley, D. L. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* **15**, 1783–1795 (2001).
- Hirose, Y., Tacke, R. & Manley, J. L. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **13**, 1234–1239 (1999).
- Kim, J. B., Yamaguchi, Y., Wada, T., Handa, H. & Sharp, P. A. Tat-SF1 protein associates with RAP30 and human SPT5 proteins. *Mol. Cell Biol.* **19**, 5960–5968 (1999).
- Li, X. Y. & Green, M. R. The HIV-1 Tat cellular coactivator Tat-SF1 is a general transcription elongation factor. *Genes Dev.* **12**, 2992–2996 (1998).
- Zhou, Q. & Sharp, P. A. Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* **274**, 605–610 (1996).
- Parada, C. A. & Roeder, R. G. A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. *EMBO J.* **18**, 3688–3701 (1999).
- Price, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell Biol.* **20**, 2629–2634 (2000).
- Ping, Y. H. & Rana, T. M. Tat-associated kinase (P-TEFb): a component of transcription preinitiation and elongation complexes. *J. Biol. Chem.* **274**, 7399–7404 (1999).
- Zhou, M. et al. Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Mol. Cell Biol.* **20**, 5077–5086 (2000).
- Dahmus, M. E. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* **271**, 19009–19012 (1996).
- Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451–462 (1998).
- Jones, K. A. Taking a new TAK on tat transactivation. *Genes Dev.* **11**, 2593–2599 (1997).

16. Garber, M. E. *et al.* The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* **12**, 3512–3527 (1998).
17. Peng, J., Zhu, Y., Milton, J. T. & Price, D. H. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.* **12**, 755–762 (1998).
18. Fong, Y. W. & Zhou, Q. Relief of two built-in autoinhibitory mechanisms in P-TEFb is required for assembly of a multicomponent transcription elongation complex at the human immunodeficiency virus type 1 promoter. *Mol. Cell. Biol.* **20**, 5897–5907 (2000).
19. Zhou, Q. & Sharp, P. A. Novel mechanism and factor for regulation by HIV-1 Tat. *EMBO J.* **14**, 321–328 (1995).
20. Marciniak, R. A. & Sharp, P. A. HIV-1 Tat protein promotes formation of more-processive elongation complexes. *EMBO J.* **10**, 4189–4196 (1991).
21. Yan, D. *et al.* CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol. Cell. Biol.* **18**, 5000–5009 (1998).
22. Perriman, R. & Ares, M. Jr ATP can be dispensable for prespliceosome formation in yeast. *Genes Dev.* **14**, 97–107 (2000).
23. Krainer, A. R. Pre-mRNA splicing by complementation with purified human U1, U2, U4/U6 and U5 snRNPs. *Nucleic Acids Res.* **16**, 9415–9429 (1988).
24. Bach, M., Bringmann, P. & Lührmann, R. Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol.* **181**, 232–257 (1990).
25. Blencowe, B. J. & Lamond, A. I. Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* **118**, 275–287 (1999).
26. Schnapp, G., Rodi, H. P., Rettig, W. J., Schnapp, A. & Damm, K. One-step affinity purification protocol for human telomerase. *Nucleic Acids Res.* **26**, 3311–3313 (1998).
27. Krainer, A. R. & Maniatis, T. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42**, 725–736 (1985).
28. Solnick, D. Alternative splicing caused by RNA secondary structure. *Cell* **43**, 667–676 (1985).
29. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell* **35**, 101–107 (1983).
30. Ares, M. Jr, Grate, L. & Pauling, M. H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **9**, 1138–1139 (1999).

Acknowledgements

We thank B. Blencowe for the gift of antisense RNA oligonucleotide and advice on the depletion procedures; E. Labourier, D. Rio, J. Underwood, D. Black and H. Wu for reagents; E. Logue and L. Zhou for technical assistance; and K. Luo, I. von Reis and S. Stroschein for discussions. Supported by grants from the National Institutes of Health to Q.Z. and the University of California Universitywide AIDS Research Program to Y.W.F.

Correspondence and requests for materials should be addressed to Q.Z.
(e-mail: qzhou@uclink4.berkeley.edu).

Crystal structure of an Eph receptor–ephrin complex

Juha-Pekka Himanen*, Kanagalaghatta R. Rajashankar†, Martin Lackmann‡, Chad A. Cowan§, Mark Henkemeyer§ & Dimitar B. Nikolov*

* Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

† Brookhaven National Laboratory, Brookhaven, New York 11973, USA

‡ Ludwig Institute for Cancer Research, Royal Melbourne Hospital, Victoria 3050, Australia

§ Center for Developmental Biology and Kent Waldrep Foundation Center for Basic Research on Nerve Growth and Regeneration, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9133, USA

The Eph family of receptor tyrosine kinases and their membrane-anchored ephrin ligands are important in regulating cell–cell interactions as they initiate a unique bidirectional signal transduction cascade whereby information is communicated into both the Eph-expressing and the ephrin-expressing cells. Initially identified as regulators of axon pathfinding and neuronal cell migration, Ephs and ephrins are now known to have roles in many other cell–cell interactions, including those of vascular endothelial cells and specialized epithelia^{1,2}. Here we report the crystal structure of the complex formed between EphB2 and ephrin-B2, determined at 2.7 Å resolution. Each Eph receptor binds an ephrin

ligand through an expansive dimerization interface dominated by the insertion of an extended ephrin loop into a channel at the surface of the receptor. Two Eph–Ephrin dimers then join to form a tetramer, in which each ligand interacts with two receptors and each receptor interacts with two ligands. The Eph and ephrin molecules are precisely positioned and orientated in these complexes, promoting higher-order clustering and the initiation of bidirectional signalling.

The communication of biochemical signals between cells is essential for the development and existence of multicellular organisms. A key method of communication is based on direct protein–protein interactions between ligands, which carry the signal, and cell-surface receptors, which recognize and transduce the information into the receiving cell. One large group of receptors and ligands, the Eph/ephrin family, sends information bidirectionally into both the receptor-expressing cell and the ligand-expressing cell^{3–6}. Ephs and ephrins are divided into two subclasses, A and B, according to their affinities for each other, sequence conservation, and the mode of ephrin membrane attachment (refs 7, 8; and Fig. 1). In general, the EphA receptors (EphA1 to EphA8) promiscuously bind to and are activated by glycosyl phosphatidylinositol (GPI)-anchored A-ephrins (ephrin-A1 to ephrin-A5), whereas EphB receptors (EphB1 to EphB6, and in some cases EphA4) bind to and are activated by the transmembrane B-ephrins (ephrin-B1 to ephrin-B3).

The membrane attachment of both Ephs and ephrins provides a mechanism whereby interactions between receptors and ligands occur only at sites of cell–cell contact, leading to the multimerization of both molecules to distinct clusters in their respective plasma membranes^{9,10}. This results in the formation of discrete bidirectional signalling centres in which the Eph receptor tyrosine kinase domain transduces the forward signal into its cell and the ephrin transduces the reverse signal into its cell. Although crystal structures for the isolated EphB2 and ephrin-B2 extracellular domains have been described^{11,12}, the precise molecular determinants that govern subclass specificity, higher-order clustering, and bidirectional signalling have not been identified. We therefore generated co-crystals and determined the structure of ephrin-B2 bound to EphB2.

The amino-terminal ligand-binding globular domain of EphB2 (refs 11, 13, 14) and the complete extracellular domain of ephrin-B2 were expressed in *Escherichia coli* and crystallized (see Methods). The structure was determined using multiwavelength anomalous diffraction¹⁵ (see Supplementary Information), and is refined at 2.7 Å resolution.

Biophysical experiments indicate that the extracellular domains of Eph receptors and ephrins interact by a multiple-step process (refs 14, 16 and 17, and J.P.H. and D.B.N., manuscript in preparation). Ephs and ephrins first bind with high affinity and specificity to form heterodimers, which are the predominant form of the complex in solution throughout a large concentration range from 10 nM to 20 μM. At higher concentrations, the dimer pairs associate into tetramers and higher-order aggregates.

The dimeric and tetrameric structures are both present in our crystals. The tetrameric complex forms a ring-like structure, in which each ligand interacts with two receptors and each receptor with two ligands (Fig. 2a, b). One of the ligand–receptor interfaces is very extensive and presumably mediates the initial high-affinity 1:1 dimerization measured in solution (arrows in Fig. 2a). The second interface is smaller (Fig. 2a) and may be responsible for the assembly of functional tetrameric ligand–receptor complexes. The carboxy termini of both ephrin-B2 molecules are located at one side of the tetramer, whereas the carboxyl termini of both EphB2 molecules are located at the other side (Fig. 2b). Such a structure is consistent with the expected positioning and orientation of both the ligand and receptor molecules on the surfaces of adjacent, juxtaposed cells.

The overall structure of ephrin-B2 in the complex (Fig. 3a, b) is

very similar to the structure of unbound ephrin-B2 (ref. 12). The ephrin-B2 folding topology is a variation on the common Greek key β -barrel fold and shares considerable similarity with the cupredoxin family of copper-binding proteins, although ephrin-B2 does not bind copper and no metals are seen in our structure. A main difference between ephrin and other cupredoxin-fold proteins is the unusual length of the ephrin $G-H_L$ and $C-D_L$ loops (where subscript L or R indicates ligand or receptor, respectively), which are part of the dimerization and tetramerization ligand-receptor interfaces, respectively (Fig. 1). These loops also undergo slight conformational readjustments on receptor binding (Fig. 3b). In the crystals of uncomplexed ephrin-B2, the molecule forms homodimers by burying hydrophobic surface regions around the $G-H_L$ loop¹². As our co-crystal analysis indicates that this loop is involved in receptor binding, it is likely that the ephrin molecules will exhibit significant rearrangements when the homodimer is displaced after interactions with EphB2.

The overall structure of the ligand-binding domain of EphB2 in the complex (Fig. 3c) is similar to that of the unbound receptor¹¹. The significant conformational changes in the receptor involve

loops at the dimerization interface. Most marked is the reorganization of the $D-E_R$ and $J-K_R$ loops, which are partially disordered in the unbound form, but become ordered to form the ligand-binding channel. Moreover, the $J-K_R$ loop has formed a short 3_{10} helix and the long disulphide-stabilized $G-H_R$ loop has moved 5 Å towards the bound ligand. The opposite end of the dimerization interface is also altered, with the short 3_{10} helix (B_R) of the unbound receptor forming a coil and the adjacent $B-C_R$ loop having undergone conformational adjustments.

Considering the EphB2-ephrin-B2 dimer (Fig. 4a), the ligand binds along the upper convex part of the β -sandwich of the receptor. The dimerization interface buries 1,160 Å² from the total receptor surface area of 9,450 Å², and 1,290 Å² from the total ligand surface of 8,140 Å². The interface can be divided into two regions. The first one encompasses the area of contact at the ephrin-docking site along the upper convex surface of the receptor and is mostly polar, relying on hydrogen-bond networks at solvent-excluded regions for binding and recognition. The interacting residues at the surface of the ephrin-B2 β -sandwich are mostly from strand G_L , with a smaller number from strands C_L and F_L . They interact with three, orthog-

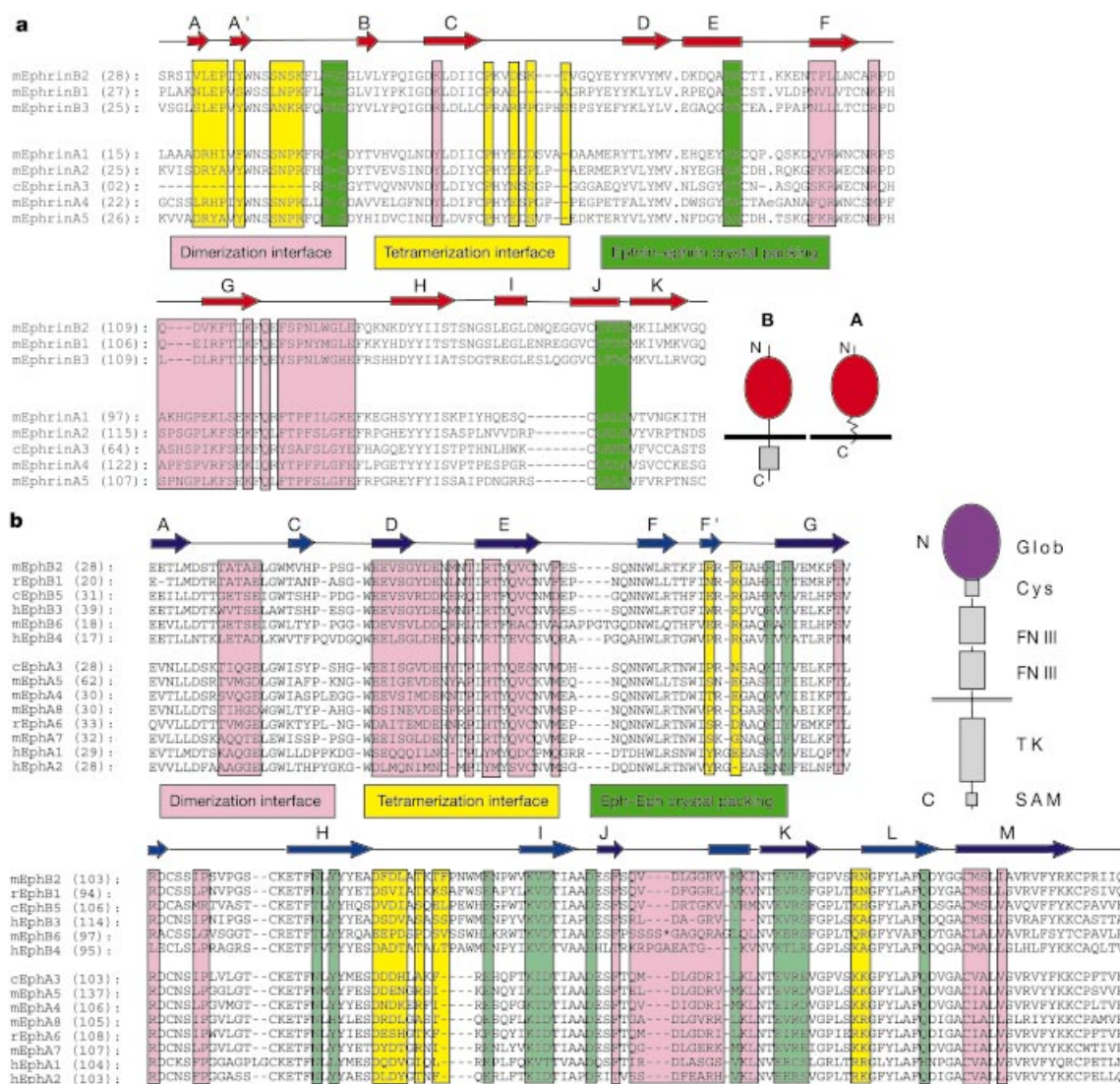


Figure 1 Sequence alignments. **a**, The receptor-binding domains of ephrins. **b**, The ligand-binding domains of Eph receptors. Secondary structural elements are depicted as arrows for β -strands and rectangles for helices. Residues in the dimerization interface are highlighted in pink, those in the tetramerization interface in yellow, and those involved in receptor-receptor or ligand-ligand crystal-packing interactions in green. Ephs are

glycosylated only outside the ligand-binding domain, whereas most ephrins have a potential glycosylation site at position Asn 39, adjacent to the tetramerization interface. Cys, cysteine-rich linker; FN III, fibronectin type III; Glob, globular domain; SAM, sterile α -motif; TK, tyrosine kinase.

onal to them and parallel to each other, layers of receptor residues from the B–C_R loop (level 1), strand E_R (level 2) and strand D_R (level 3), respectively (see Figs 1 and 4a). In each of these layers, hydrogen bonds connect backbone atoms from one of the molecules to side chains from the other, including Asp110_L–Thr38_R (level 1), Leu101_L–Thr263_R (level 2) and Lys112_L–Ser55_R (level 3). Two of the layers also contain backbone–backbone hydrogen bonds, including Thr38_R–Lys112_L (level 1) and Ser55_R–Thr114_L (level 3). Side-chain hydrogen bonds (Gln68_R–Gln118_L), and salt bridges (Glu40_R–Lys112_L and Glu52_R–Lys116_L) complete the interaction network.

The second region of the dimerization interface centres around the G–H_L loop of ephrin-B2, which is inserted in a channel on the surface of EphB2 (Fig. 4b, c). Strands D_R, E_R and J_R define the two sides of the channel, and strands G_R and M_R line its back. The ligand binds by approximating the side of its β-sandwich to the outside surface of the channel. Specifically, strands F_L, G_L and C_L dock against strands D_R and E_R of EphB2. The ligand inserts its long G–H_L loop into the channel, which becomes buttressed by the

receptor's G–H_R loop closing-in from the top of the β-sandwich. The binding here is dominated by van der Waals contacts between two predominantly hydrophobic surfaces as the ligand buries Phe 120_L, Pro 122_L, Leu 124_L, Trp 125_L and Leu 127_L. Of these residues, Phe 120_L is the only one that does not interact with the sides of the channel, but instead with EphB2 residues Ile 108_R and Pro 109_R from the long G–H_R loop at the top of the binding interface. At the tip of the inserted ephrin loop are Leu 124_L and Trp 125_L, with the side chain of Trp 125_L reaching to the very end of the channel. Pro 122_L is in direct contact with the Cys70_R–Cys192_R disulphide bridge. In addition to the many van der Waals interactions, there are several intermolecular hydrogen bonds involving main-chain atoms, including Phe120_L–Ser101_R and Pro122_L–Arg103_R.

Like the dimerization interface, the tetramerization interface can also be divided into two regions involving different sets of secondary structural elements (Fig. 4d). The first region is located on the top of the structure and contains most of the ligand–receptor contacts, including the A–A' β-strand of ephrin-B2 and residues from the

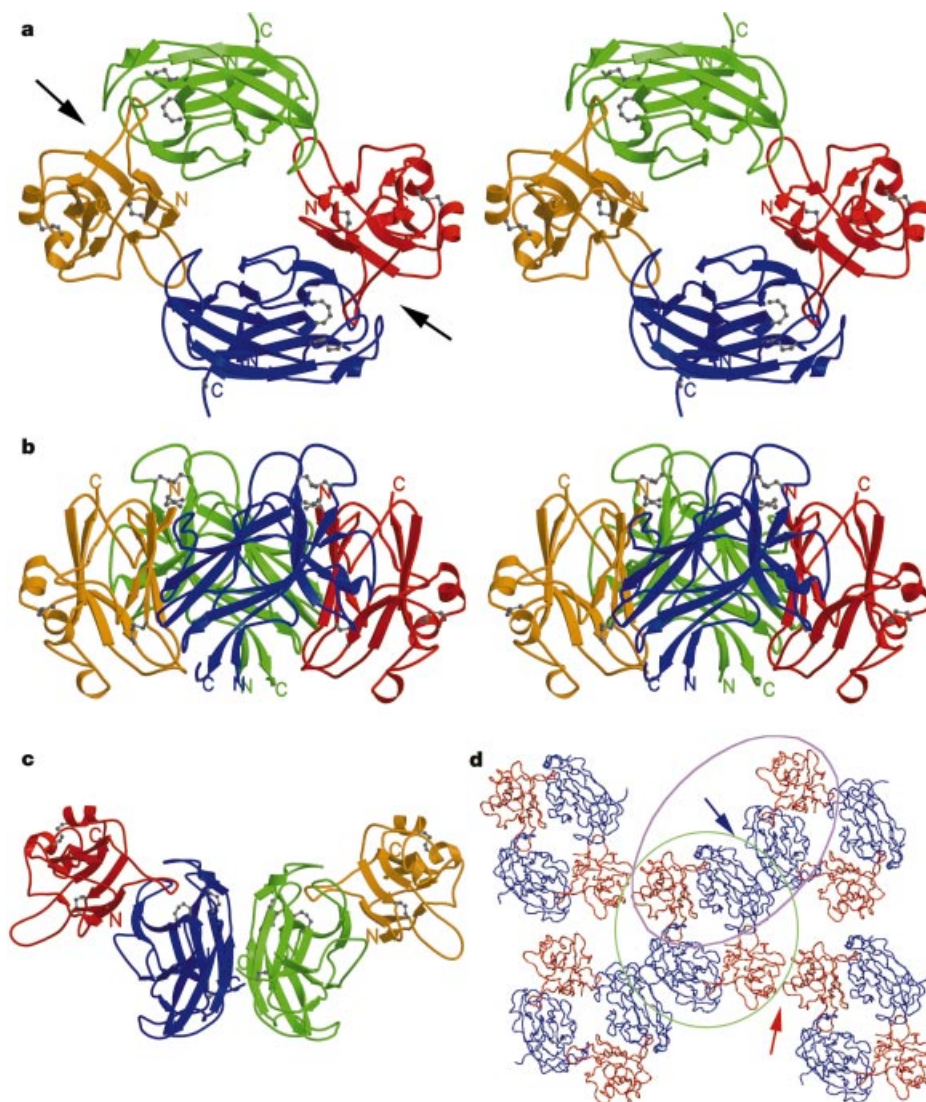


Figure 2 Structure of the EphB2–ephrin-B2 tetramer. **a, b**, Two orthogonal stereo views of the EphB2–ephrin-B2 tetramer. Ligand molecules are in red and orange, and the receptors in blue and green. The tetramer is generated by two ligand–receptor dimers (blue–red and green–orange) joining together. Ligand–receptor dimerization interfaces are indicated by arrows in **a, c**. Two EphB2 molecules joined by a receptor–receptor

interface observed in the crystals, and each bound to a single ligand molecule. **d**, Arrangement of the EphB2 (blue) and ephrin (red) molecules in the crystal. There are four ligand and four receptor molecules in the asymmetric unit. Two possible tetramers are indicated with green or purple lines. The ligand–ligand (red arrow) and receptor–receptor (blue arrow) crystal-packing interfaces are indicated.

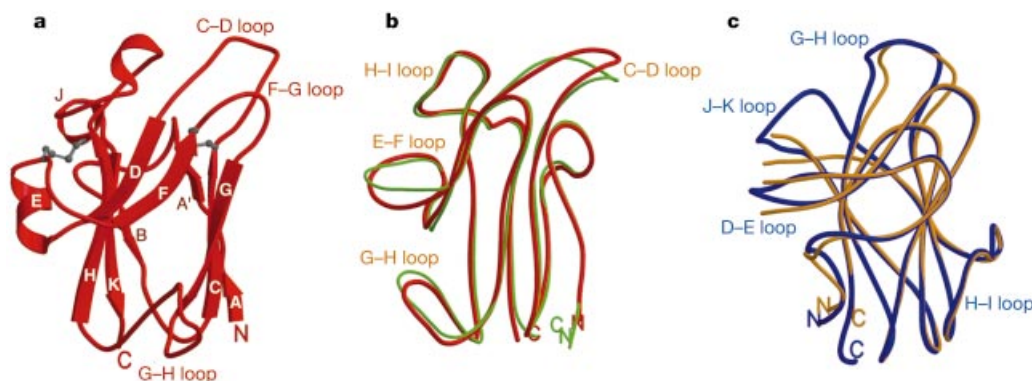


Figure 3 Structure of ephrin-B2 and conformational changes induced in ephrin-B2 and EphB2 on complex formation. **a**, Structure of the extracellular receptor-binding domain of ephrin-B2. Secondary structural elements are indicated and disulphide bonds are shown in grey. **b**, Receptor-bound ephrin-B2 (red) superimposed on the structure of the unbound ligand (green). The free and bound ligands are superimposable with a root mean-square

deviation (r.m.s.d.) of 1.3 Å between Cα positions. **c**, Ligand-bound EphB2 (blue) superimposed on the structure of the unbound receptor (orange). The strands of the β-sandwich are superimposable with an r.m.s.d. of 0.6 Å between Cα positions. The D-F_R and J-K_R loops of the receptor become structured on binding to ephrin-B2, whereas the G-H_R loop moves towards the bound ligand.

long, H-I_R specificity loop of the receptor¹¹. In the second region, the long C-D_L loop of the ligand has closed towards the receptor and interacts with the F'-G_R and K-L_R loops. In total, the tetramerization interface is relatively small, burying 500 Å² of receptor and 500 Å² of ligand surface area, and is mostly polar. There are few hydrogen bonds between backbone and side-chain atoms, but several weak hydrogen bonds between only side-chain atoms. In addition, Arg 84_R forms a salt bridge with Asp 69_L. A small number of van der Waals contacts are also present, the most predominant of which is the stacking of the aromatic rings of Phe 128_R and Tyr 37_L.

In our crystals, there is a possible alternative tetramer formed by the back-to-back joining of EphB2 molecules (Fig. 2c). However, the idea that the circular ligand-receptor tetramer shown in Fig. 2a, b, and not the tetramer shown in Fig. 2c, is the biologically relevant complex is supported by mutagenesis data (ref. 11; and Andrew Boyd and M.L., unpublished data), by the size of the area buried on tetramerization (circular tetramer, 2,000 Å²; tetramer on Fig. 2c, 1,100 Å²), and by the absence of the back-to-back receptor-receptor interface in the crystals of the unbound EphB2 protein¹¹.

In our crystals the tetramers pack through receptor-receptor (burying 550 Å² in each molecule) and ligand-ligand (burying 350 Å² in each molecule) interfaces (Figs 1, 2c, d). The absence of these interfaces in crystals of the unbound proteins^{11,12} suggests that these interactions become energetically favourable after heterodimerization and subsequent tetramerization of the Eph and ephrin proteins, and may help to explain the higher-order clustering observed for Ephs and ephrins.

Finally, the dimer and to a lesser extent tetramer interfaces observed in the crystals provide insight into the basis of Eph/ephrin subclass specificity. B-ephrins contain a bulky polar residue at position 109_L (glutamine or leucine), which our structure shows to be hydrogen bonded to the conserved Thr 38_R found in the EphB-subclass receptors. In EphA-subclass receptors, position 38_R is occupied by a bulky residue (usually glutamine or methionine), whereas the A-ephrin ligands have serine or alanine at position 109_L. Such an arrangement would maintain a favourable contact in either A- or B-subclass complexes, but not in mixed A+B complexes. In a similar way, Thr 114_L is in van der Waals contact with Val 54_R, and these residues are conserved in the B-subclass molecules. In the A-subclass receptors, the corresponding 54_R position is occupied by a residue with a larger hydrophobic side chain (isoleucine or methionine), whereas its partner at position 114_L of A-ephrins, a serine residue, has a smaller side chain. Notably, EphA4, which has been shown to bind some B-ephrins, is the only A-subclass receptor that contains a valine at position 54_R. In addition, the sequence of the

G-H_L loop of ephrins is different in the A and B subclasses. For example, the side chain of ephrin-B2 Asn 123_L is only 3 Å away from the side chain of EphB2 Ser 194_R. These polar residues are conserved in the B-subclass molecules, but are replaced by the non-polar Phe 123_L and Ala 194_R in all A-subclass molecules. Furthermore, the side chains of residues Leu 127_L, Phe 113_L and Tyr 57_R are close together in the B-subclass complexes and are involved in van der Waals contacts, including an aromatic-aromatic interaction. In EphA-subclass receptors, position 57_R is occupied by a large hydrophobic, non-aromatic amino acid, whereas there is an aromatic amino acid at position 127_L in A-ephrins. Therefore, A-subclass complexes would most probably maintain a similar aromatic-aromatic contact in this region, but with a switch in the position of one of the rings. By contrast, mixed (A+B)-subclass complexes would have to approximate an odd number of aromatic side chains, which might be energetically less favourable.

Combined with previous reports^{11,12}, our structural studies suggest the following mechanism for the initiation of Eph/ephrin signalling. Before cell-cell contact, the Eph and ephrin molecules are loosely pre-clustered at distinct locations on the surface of their cell, most probably in cholesterol-rich lipid rafts¹⁸. Unbound ephrins may form low-affinity ephrin-ephrin homodimers¹². On cell-cell contact, the Eph receptor binds its ephrin ligand with 1:1 stoichiometry and nanomolar affinity through the extensive heterodimerization interface. As the Eph-ephrin heterodimerization interface overlaps with the proposed ephrin-ephrin homodimer interface, the ligand molecules become reoriented after receptor binding. Eph-ephrin heterodimerization then creates complementary interaction surfaces that result in the joining of dimer pairs into the tetrameric complexes (Fig 2a, b). These circular tetramers are very stable and fix the orientation of the participating ligands and receptors, thereby allowing the tyrosine kinase domains of the paired receptor molecules to *trans*-autophosphorylate each other and initiate forward signalling. Notably, the disruption of ephrin-ephrin homodimers on Eph binding and the formation of the circular tetramers are both likely to result in a repositioning of the ephrin transmembrane and cytoplasmic domains, perhaps converting them from an inactive or closed configuration to an active configuration. This would allow tyrosine phosphorylation of the ephrin cytoplasmic tail and the initiation of reverse signalling.

Our structural analysis further indicates that the circular tetramers can be arranged in aggregates by means of weaker receptor-receptor and ligand-ligand interfaces, providing mechanistic insight into the higher-order clustering observed for these

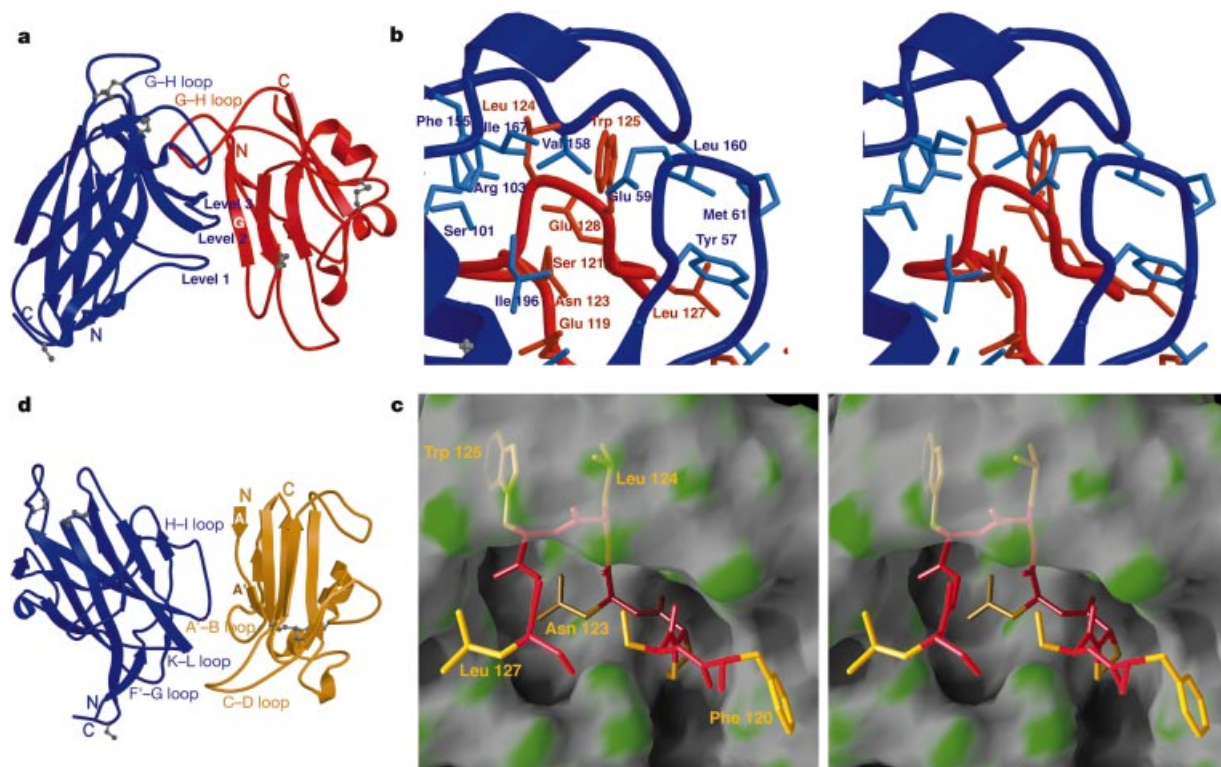


Figure 4 Structures of the EphB2–ephrin-B2 dimer and tetramer interfaces. **a**, The ligand–receptor dimer interface with EphB2 in blue and ephrin-B2 in red. **b**, The channel on the surface of EphB2 viewed in stereo from the distal end of the ligand, showing the receptor (blue) and the incoming ephrin loop (red). Only the tip of the side chain of Trp 125_L reaches the end of the channel. **c**, Molecular surface representation of EphB2,

viewed from position of the incoming ephrin, showing the entrance of the hydrophobic channel with the deeply inserted ephrin-B2 loop (in stereo). The molecular surface of the receptor is colour-coded according to curvature with concave regions in grey and convex regions in green. **d**, The ligand–receptor tetramerization interface with EphB2 in blue and ephrin-B2 in orange.

molecules. In addition to the interaction domains described here, other regions of the Eph and ephrin are probably also involved in the final positioning of bidirectional signalling complexes, including the extracellular cysteine-rich linker¹⁴ and the intracellular SAM domain of the receptors^{19,20}, and the C-terminal PDZ-domain-binding sites found in both receptors and ligands^{21–25}.

This structure of an Eph–ephrin complex explains many of the biochemical and signalling properties of these molecules. Of particular note, our structure gives a framework for understanding the basis for the different binding affinities that bring about the A- and B-subclass specificities. Furthermore, it provides detail into how the juxtaposition of ligand-expressing and receptor-expressing cells might lead to formation of ordered receptor–ligand multimeric clusters at cell–cell interfaces. Given the potential importance of Ephs and ephrins in cardiovascular function, nerve regeneration and cancer, these structures may yield insight into the future development of therapeutic agents. □

Methods

Protein expression

Expression and purification of murine EphB2 globular domain (residues 28–210) and the extracellular domain of murine ephrin-B2 (residues 25–233) was done as described¹¹. For the production of seleno-methionine-labelled EphB2 protein, the expression vector was transformed into the methionine auxotroph *Escherichia coli* B834 and grown in the presence of 50 mg l⁻¹ seleno-methionine. We removed the histidine-containing amino-terminal thioredoxin fusion sequences by thrombin proteolysis. After formation, the ligand–receptor complex was subjected to further limited proteolysis with proteinase V8 (Sigma), which removed the non-structured C-terminal region of ephrin-B2 and some of the remaining amino-terminal vector-derived sequences. We purified the resulting complex, containing ephrin-B2 (25–187) and EphB2 (28–210) plus eight N-terminal amino acids in each molecule from the expression vector, by anion-exchange and gel-

filtration chromatography. DNA sequencing of plasmids and mass spectrometry showed that the proteins used for crystallization were not modified by their expression and purification. Binding studies, done as described¹¹, showed that EphB2 and ephrin-B2 form a high-affinity heterodimeric complex with a *K*_d of 25 nM.

Crystallization

The purified EphB2–ephrin-B2 complex was concentrated to 20 mg ml⁻¹ in a buffer containing 10 mM KCl, 2 mM MgCl₂ and 10 mM HEPES (pH 7.2), and crystallized in a hanging drop by vapour diffusion at 20 °C against a reservoir containing 1.6 M sodium/potassium-phosphate and 100 mM HEPES-NaOH (pH 7.5) (Hampton Research). Crystals grew in the *P*1 space group (*a* = 78 Å, *b* = 78 Å, *c* = 78 Å; *a* = 69°, *b* = 75°, *c* = 69°) with four receptors and four ligands (two tetramers) in the asymmetric unit.

Data collection and structure determination

We collected data either in-house using a Rigaku RAXIS-IV imaging plate area detector at the NLSL Brookhaven beamlines X25 and X9A, or at CHESS beamline F2. Images were integrated, scaled and merged using DENZO and SCALEPACK²⁶. Subsequent calculations were performed with the CCP4 program suite²⁷. The structure of the EphB2–ephrin-B2 complex was determined using multiple wavelength anomalous diffraction phasing¹⁵ (see Supplementary Information). Peak wavelength anomalous data were input to the program SnB (ref. 11; and J.P.H. and D.B.N., manuscript in preparation) to identify the location of the selenium atoms. We refined the peaks using MLPHARE²⁷ in the resolution range 25–3.0 Å. The calculated phases had a figure of merit of 0.36, which was improved to 0.83 by density modification procedures with the program DM²⁷. Fourfold non-crystallographic symmetry averaging resulted in substantial improvement of the electron density. The receptor molecules were manually placed in the map, and ligand molecules were built using the program O (ref. 28). Refinement of the model by conventional least-squares algorithm was done with X-PLOR²⁹.

The final refined model at 2.7 Å resolution has a free *R* value of 29.5% with restrained refinement of temperature factors. Several residues in the G–H_R and E–F_R loops of EphB2 and the C–D_L loop of ephrin-B2 are partially disordered, but are nevertheless built into the model, which includes residues 28–207 of the receptor and residues 30–169 of the ligand. Stereochemical analysis of the refined models using PROCHECK²⁷ revealed main-chain and side-chain parameters better than, or within, the typical range of values for protein structures determined at corresponding resolution (overall *G* factor = 0.1).

Received 10 July; accepted 15 October 2001.

1. Flanagan, J. G. & Vanderhaeghen, P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurosci.* **21**, 309–345 (1998).
2. Frisen, J., Holmberg, J. & Barbacid, M. Ephrins and their Eph receptors: multitasking directors of embryonic development. *EMBO J.* **18**, 5159–5165 (1999).
3. Henkemeyer, M. *et al.* Nuk controls pathfinding of commissural axons in the mammalian central nervous system. *Cell* **86**, 35–46 (1996).
4. Holland, S. J. *et al.* Bidirectional signalling through the EPH-family receptor Nuk and its transmembrane ligands. *Nature* **383**, 722–725 (1996).
5. Bruckner, K., Pasquale, E. B. & Klein, R. Tyrosine phosphorylation of transmembrane ligands for Eph receptors. *Science* **275**, 1640–1643 (1997).
6. Cowan, C. A. & Henkemeyer, M. The SH2/SH3 domain adaptor Grb4 transduces B-ephrin reverse signals. *Nature* **413**, 174–179 (2001).
7. Eph Nomenclature Committee. Unified nomenclature for Eph family receptors and their ligands, the Ephrins. *Cell* **90**, 403–404 (1997).
8. Gale, N. W. *et al.* Eph receptors and ligands comprise two major specificity subclasses and are reciprocally compartmentalized during embryogenesis. *Neuron* **17**, 9–19 (1996).
9. Henkemeyer, M. *et al.* Immunolocalization of the Nuk receptor tyrosine kinase suggests roles in segmental patterning of the brain and axonogenesis. *Oncogene* **9**, 1001–1014 (1994).
10. Davis, S. *et al.* Ligands for EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity. *Science* **266**, 816–819 (1994).
11. Himanen, J. P., Henkemeyer, M. & Nikolov, D. B. Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2. *Nature* **396**, 486–491 (1998).
12. Toth, J. *et al.* Crystal structure of an ephrin ectodomain. *Dev. Cell* **1**, 83–92 (2001).
13. Labrador, J. P., Brambilla, R. & Klein, R. The N-terminal globular domain of Eph receptors is sufficient for ligand binding and receptor signaling. *EMBO J.* **16**, 3889–3897 (1997).
14. Lackmann, M. *et al.* Distinct subdomains of the EphA3 receptor mediate ligand binding and receptor dimerization. *J. Biol. Chem.* **273**, 20228–20237 (1998).
15. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
16. Lackmann, M. *et al.* Ligand for EPH-related kinase (LERK) 7 is the preferred high affinity ligand for the HEK receptor. *J. Biol. Chem.* **272**, 16521–16530 (1997).
17. Stein, E. *et al.* Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses. *Genes Dev.* **12**, 667–678 (1998).
18. Bruckner, K. *et al.* EphrinB ligands recruit GRIP family PDZ adaptor proteins into raft membrane microdomains. *Neuron* **22**, 511–524 (1999).
19. Stapleton, D., Balan, I., Pawson, T. & Sicheri, F. The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nature Struct. Biol.* **6**, 44–49 (1999).
20. Thanos, C. D., Goodwill, K. E. & Bowie, J. U. Oligomeric structure of the human EphB2 receptor SAM domain. *Science* **283**, 833–836 (1999).
21. Hock, B. *et al.* PDZ-domain-mediated interaction of the Eph-related receptor tyrosine kinase EphB3 and the ras-binding protein AF6 depends on the kinase activity of the receptor. *Proc. Natl Acad. Sci. USA* **95**, 9779–9784 (1998).
22. Torres, R. *et al.* PDZ proteins bind, cluster, and synaptically colocalize with Eph receptors and their ephrin ligands. *Neuron* **21**, 1453–1463 (1998).
23. Buchert, M. *et al.* The junction-associated protein AF-6 interacts and clusters with specific Eph receptor tyrosine kinases at specialized sites of cell–cell contact in the brain. *J. Cell Biol.* **144**, 361–371 (1999).
24. Lin, D., Gish, G. D., Songyang, Z. & Pawson, T. The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif. *J. Biol. Chem.* **274**, 3726–3733 (1999).
25. Cowan, C. A., Yokoyama, N., Bianchi, L. M., Henkemeyer, M. & Fritzsche, B. EphB2 guides axons at the midline and is necessary for normal vestibular function. *Neuron* **26**, 417–430 (2000).
26. Otwinowski, Z. & Minor, W. *Data Collection and Processing* 556–562 (SERC Daresbury Laboratory, Warrington, UK, 1993).
27. Project, C. C. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, I. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Brünger, A. T. *X-PLOR v. 3.1 Manual* (Yale Univ. Press, New Haven, 1993).

Supplementary information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Coyle for help with figure preparation, and P. Jeffrey for help with X-ray measurements. D.B.N. is a PEW fellow. This work was supported by the NIH and the New York Council Speaker's Fund for Biomedical Research (D.B.N.) and by the Welch Foundation (M.H.).

Correspondence and requests for materials should be addressed to D.B.N. (e-mail: dimitar@ximpact3.ski.mskcc.org). Coordinates are deposited in the Brookhaven Protein Data Bank (accession code 1KGY).

correction

Neurogenesis in the adult is involved in the formation of trace memories

Tracey J. Shors, George Miesegaes, Anna Beylin, Mingrui Zhao, Tracy Rydel & Elizabeth Gould

Nature **410**, 372–376 (2001).

The original Fig. 2e–j is in error because it was based on incorrectly sorted brain tissue. A corrected figure is shown below and demonstrates that the anti-mitotic agent MAM does not affect the gross morphology or the volume of either the hippocampus or the cerebellum. The revised figure is based on a new set of tissue obtained from brains in experiment 3 that had been exposed to MAM ($n = 3$) or saline ($n = 3$) for 14 days and processed as described. Sections from each of these brains are presented. This error (and its correction) does not alter the main findings we reported, namely that treatment with MAM depletes newly generated neurons in the hippocampus and impedes learning of trace eyeblink conditioning in the adult rat. □

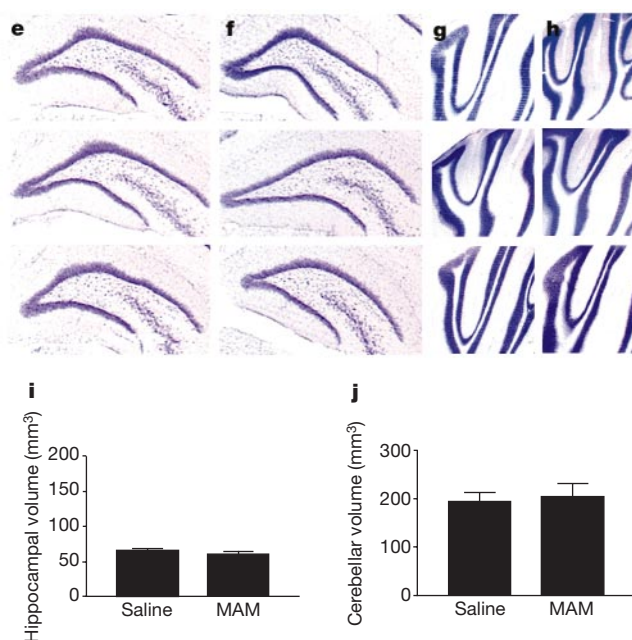


Figure 2e–j MAM treatment does not alter the gross morphology of the hippocampus or cerebellum. **e, f**, Cresyl violet stained sections of the dentate gyrus from three saline-treated (**e**) and three MAM-treated (**f**) adult rats. **g, h**, Cresyl violet stained sections of the cerebellar cortex from three saline-treated (**g**) and three MAM-treated (**h**) adult rats. **i, j**, Hippocampal (**i**) and cerebellar (**j**) volume (mean $\pm$ s.e.m.).

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:
Leonie Weiss (4954)

European Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Sille Opstrup (4994)

Spain/ Portugal: Leonie Weiss
(4954)

Production Manager: Billie Franklin

To send films and materials use
London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Backing a meeting of minds

Among countless other things, the events of 11 September made scientists reconsider the importance of travel and conferences. After the crashes in New York, Washington and Pennsylvania, some meetings were postponed or cancelled and, for a moment, the future of scientific communication looked uncertain. But many scientists soon grew defiant, emphasizing that such setbacks were only temporary.

"Many of my best ideas are germinated by insights that come from face-to-face discussions. I, for one, cannot and will not allow this week's events to change that," John Quackenbush, a faculty member at The Institute for Genomic Research in Rockville, Maryland, told me shortly afterwards.

Quackenbush's reason was as eloquent as his rallying cry. "We are a community, something that to a large extent transcends boundaries and ideologies," he said. "While we may disagree with each other, I believe that most of us still operate in an atmosphere of mutual respect." Although technological tools help to tie that community together, they don't replace live presentations and the ensuing discussions. In fact, electronic exchanges often prompt people to share their ideas, results and progress in the flesh.

There also seems to be less chance for serendipity in virtual reality. Some of the most vital moments of a conference occur during the buzz after a keynote speech, overheard conversations about proposed initiatives or chance encounters at receptions. And a good conference can leave one with the palpable sense of where a particular field is heading.

Naturejobs is this week pleased to offer its 2002 events directory in both print and electronic form. We hope that it will inform and encourage meeting attendance in the coming year.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Immunologist returns to Austria to head new institute; Structural GenomiX appoints scientific officer; mathematician leaves Hewlett-Packard for Harvard; and more [Back page](#)

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

BIOLOGY

Immunologist **Josef Penninger** has been lured back to his native Austria to set up and run a new institute. Penninger, a member of the Amgen Institute in Toronto since 1994, this month formally agreed to head the Institute for Molecular Biotechnology of the Austrian Academy of Sciences.

Penninger was initially hesitant about moving away from the tight focus of his research in cell signalling. But he eventually decided that being presented with a “blank slate” was a rare chance to gain complete academic freedom. “I can decide what research direction to take, what researchers to hire,” Penninger says. He has already offered input on the new building’s design. Construction on the site near downtown Vienna will begin early in 2002. Penninger will stay with Amgen until next year and retain his lab there until his existing graduate students finish their projects.

BIOLOGY AND PHYSICS

Next month, **Jeremy Gunawardena** will move from Hewlett-Packard’s Basic Research Institute in the Mathematical Sciences, which he established in Bristol, UK, to become a director of the systems biology programme at Harvard University’s Bauer Center for Genomics Research. Gunawardena, who has been with Hewlett-Packard since 1987, decided to switch countries and sectors after he realized the company was not going to move aggressively into biosciences. The Harvard institute is using tools and techniques from mathematics, physics and computer sciences to tackle biological problems. Gunawardena was attracted to it because it will allow him to apply his skills in an attempt to understand complex systems. “They needed new tools and new ways of thinking about problems,” he says. He is happy to have a short-term appointment, as he is keen to keep moving towards programmes that are responding aggressively to scientific changes. “I’m not going to be at any one place for 10 years,” Gunawardena explains.

BIOTECHNOLOGY

San Diego-based biotech company Structural GenomiX will next month get a new chief scientific officer. **Stephen Burley** is leaving his post at Rockefeller University in New York after 11 years to take up the new position. At Rockefeller,

Burley directed a structural biology lab, held an endowed professorship and received a Howard Hughes Medical Institute grant. Together, those positions gave him what he calls “one of the best academic jobs in the country”. But industry beckoned.

After Structural GenomiX this spring bought Prospect Genomics, a San Francisco-based bioinformatics company he co-founded, he soon realized that the move made sense. Structural GenomiX has automated equipment for protein structure discovery, as well as access to a beamline at the Advanced Photon Source at the Argonne National Laboratory near Chicago in Illinois. That, along with the software from Prospect, will allow Burley and his colleagues to solve structures of proteins bound to small molecules. Doing so will allow the company to see how subtle variations in a protein or a drug affects the binding of the two, an essential component of rational drug design. The company expects to recruit about 30 scientists by the summer, mostly chemists.

BIOTECH MANAGEMENT

NicOx, a French biotech company based in Sophia Antipolis, has appointed **Marco Sardina** as senior medical director and **Romano Rivolta** as director of chemistry and formulation. Since 1997, Sardina has been medical director of Italfarmaco, a biotech company based in Milan, Italy. Before joining NicOx, Rivolta was technology-transfer manager at Patheon Italia, a pharmaceutical firm near Milan.

Last month **Ian Gilham** was appointed group managing director, laboratory division, at Axis-Shield, a diagnostics company based in Dundee, Scotland. Gilham joins the company from pharmaceuticals giant GlaxoSmithKline, where he was vice-president of pharmacogenetics and applied diagnostics.

Ingenium Pharmaceuticals, a biotech company in Martinsried, Germany, has appointed **Erich Platzer** and **Richard Aldrich** to its supervisory board. Platzer is chairman of Novuspharma, a drug-discovery company based in Monza, Italy, and is the former therapeutic area head of oncology global strategic marketing at drugs company Hoffmann-La Roche. Aldrich is the former senior vice-president and chief business officer of biotech firm Vertex Pharmaceuticals in Cambridge, Massachusetts.



Josef Penninger



Jeremy Gunawardena



Romano Rivolta

CONTACTING US AT MOVERS

Please send any information on new appointments or job moves to: naturejobseditor@naturedc.com.